CONTEMPORARY PHYSICS

A Journal of Interpretation and Review

Editor : G. R. Noakes, M.A., F.INST.P.

Editorial Board:

Consultant Editors: Sir Lawrence Bragg, O.B.E., M.C., M.A., D.SC., F.R.S.
Sir George Thomson, M.A., D.SC., F.R.S.

Advisory Editors: Professor L. F. Bates, PH.D., D.SC., F.R.S.
Dr. K. A. G. Mendelssohn, F.R.S.
Dr. F. A. Vick, O.B.E.

VOLUME 2
OCTOBER 1960 to AUGUST 1961.

ALERE FLAMMAM

Printed and Published by
TAYLOR & FRANCIS LTD
RED LION COURT, FLEET STREET, LONDON, E.C.4

CONTENTS OF VOL. 2. (Nos. 1–6)

NUMBER 1—OCTOBER 1960

Masers. Professor K. W. H. Stevens 1
O.E.E.C. Seminar on the Use of Television for the Teaching of Science 14
The International Scientific Radio Union... 14
The Neutron Booster. Dr. M. J. Poole 15
The Sceptical Physicist. Dr. R. A. R. Tricker 21
The Tandem Generator. Dr. W. D. Allen 31
The Physical Science Study Committee
    (3) The Planning and Structure of the Course. Mr. Stephen White 39
Physics at the British Association, Cardiff Meeting, 1960. Professor L. F. Bates 55
The Nuclear Physics Meeting of the Institute of Physics and the Physical Society, Glasgow, September, 1960. Dr. W. S. C. Williams ... 59
Opening of the Atomic Energy Establishment, Winfrith. Mr. V. T. Saunders 63
International Conference on Physics Education 69
Conference of the Physical Sciences Section of the Association of Teachers in Colleges and Departments of Education. Mr. E. J. Wenham 71
Contemporary Reading 73
Essay Review of *La Structure et l'Evolution de l'Univers* by Dr. E. H. Linfoot 75
Book Reviews 78

NUMBER 2—DECEMBER 1960

Energy for the World's Needs. Mr. L. Rotherham 81
Faraday and the Structure of Matter. Proffesor L. Pearce Williams 93
Science in Space. Dr. I. M. Stephenson 106
Corso di Aggiornamento. Mr. G. R. Noakes... 115
Radar Astronomy. Dr. J. V. Evans 116
Non-linear Vibrations. Dr. E. W. Lee 143
The Post Office Research Station. Mr. V. T. Saunders 157

NUMBER 3—FEBRUARY 1961

Advancing Fronts in Meteorology. Sir Graham Sutton 165
X-Ray Microscopy. Dr. W. C. Nixon 183
The Friction of Solids. Dr. E. H. Freitag 198
The Physics of Diamond. Dr. E. W. J. Mitchell 217
Low Temperature Physics. Dr. D. Shoenberg 223
Contemporary Reading 231
Optics at Imperial College 235
The 45th Annual Exhibition of The Institute of Physics and The Physical Society 236
The Science Masters' Association Annual Meeting 238
D.S.I.R. Grants and Awards, 1961 240
Book Reviews 241

NUMBER 4—APRIL 1961

Science in Civil Defence. Dr. J. McAULAY 245

Satellite Orbits and their Geophysical Implications. Mr. D. G. KING-HELE 253

Interference Spectroscopy. Dr. H. J. J. BRADDICK 268

The Microtron. Dr. R. E. JENNINGS 277

Electromagnetic Induction. Dr. R. R. BIRSS 286

Physical Science in the Training College. Mr. E. J. WENHAM ... 298

Thermoelectric Power Generation. Mr. P. J. BATEMAN 302

The National Chemical Laboratory Open Day. Mr. V. T. SAUNDERS 312

Mathematics for Physicists and Engineers 318

Contemporary Reading 319

Book Reviews 321

NUMBER 5—JUNE 1961

Particles, Field Theory and Force. Dr. R. J. BLIN-STOYLE 325

Electrification Following the Contact of Solids. Dr. W. R. HARPER 345

The Gyromagnetic Ratio of the Proton. Dr. P. VIGOUREUX ... 360

The Scalpel and the Spectroscope. Dr. J. M. A. LENIHAN 367

Physics for Grammar Schools. 381

On the Record 382

A Survey of Research on the Transuranic Elements. Dr. M. B. WALDRON 385

The Teaching of Modern Physics in Schools. SISTER ST. JOAN OF ARC 403

NUMBER 6—AUGUST 1961

The Applications of Single Crystals. Mr. A. E. DE BARR 409

Neutrino and Antineutrino Losses in Nuclear Processes. Dr. G. M. LEWIS 423

Radio Emission from Interstellar Neutral Hydrogen. Dr. R. D. DAVIES 428

The Nature and Properties of a Plasma. Dr. A. VON ENGEL 447

Accurate Voltage Reference Systems. Mr. G. W. BROWN 463

The Correspondence between a Current-carrying Coil and a Magnetic Dipole. Dr. J. ALAN CHALMERS 472

DSIR Research Establishments. Mr. V. T. SAUNDERS 475

Contemporary Reading 479

Essay Review of A. Abragam's *The Principles of Nuclear Magnetism* Dr. J. G. POWLES 482

Book Reviews 484

Index to Vol. 2 489

Masers

by K. W. H. STEVENS
The University, Nottingham

Summary

This article is intended to give the reader some understanding of why masers have suddenly become of such great interest. It begins with an account of the background to the problem of microwave amplification and then passes to a description of the properties of the populations of energy levels in thermal and other kinds of equilibrium. Particular attention is paid to the way in which a population reversal can be used for amplification, the basis of the maser. The next considerations are related to the performance characteristics to be anticipated and the article ends with descriptions of the ammonia maser, the ruby 3-level maser and a brief mention of the 2-level maser work. In having to choose between describing the technological and physical problems which are involved the basic physics has seemed of more importance and has been stressed.

1. Introduction

The contribution which the development of radar made to the conduct of the last war is well known. What is not perhaps so widely appreciated is that it placed in the hands of post-war physicists a powerful new tool with which to explore the properties of materials. Military requirements demanded that devices and techniques should be developed for work at very short radio wavelengths, and physicists and electronic engineers responded by producing a whole range of equipment for use at microwave frequencies. Thus when the war ended physicists were much more favourably placed than ever they had been before for working at microwave frequencies. In consequence there was a considerable burst of activity in what is now known as microwave spectroscopy. That is, spectroscopy using electromagnetic waves with wavelengths of the order of centimetres.

The device that really made the difference was the klystron, which is a tunable oscillator giving a very steady output at a definite frequency, with a power of the order of tens of milliwatts. No amplifier for microwave frequencies was developed, but this has not proved a serious limitation. The reason is that usually the information in which one is interested is carried as a modulation on the microwave signal. Rectification of microwaves is possible, using crystal rectifiers, and the rectified power fluctuates at the modulation frequency, so that it can be amplified by conventional valve amplifiers. Nevertheless it has been a challenge to develop microwave amplifiers, and since the war considerable progress has been made, particularly with the so-called travelling wave tube amplifiers. They find a use particularly in repeater stations, where it is necessary to amplify a weak incoming microwave signal and then re-radiate it towards another such repeater station.

In assessing the relative merits of amplification at the carrier or at the modulation frequency an important criterion is the signal to noise ratio. Any amplifier produces an output signal even with no input signal, the so-called noise output. This arises because no amplifier is free of resistive losses, and anything which is capable of absorbing energy will also emit energy. Thus associated with every resistive element in the amplifier there is a random voltage related to its emission of energy. The resistances in the first stages of an amplifier are responsible for a large part of the output noise, for their random voltages receive maximum amplification. However the random voltages from a resistor can be reduced by cooling, for the noise power is proportional to the absolute temperature. Thus, in principle at least, it should be possible to keep the resistive noise down to any chosen level by cooling the resistors sufficiently. However, in conventional valves, there are other sources of noise, associated with the electron flow between the electrodes. A resistive contribution comes because the electron beam presents a resistance to any external circuit coupling the electrodes, but there is another form of noise associated with the finite charge of the electron. It is not easy to reduce these latter valve noises below certain limits (for example, a hot cathode is usually needed) and where the valve noises are greater than the resistor noise at room temperature there is little to be gained by cooling the resistors. In trying to receive a weak signal it is essential that when it has passed through the amplifier the amplified signal should be recognizable above the amplifier noise. Thus whether it is better to amplify at the carrier or modulation frequency, after rectification, depends on the signal to noise ratio of the two amplifiers. The travelling wave tubes are fairly noisy, and it is usual, except in special circumstances, to amplify microwave signals at the modulation frequency. Of course if the incoming signal is a strong one the above considerations do not apply, for the amplified signal will, in any case, be well above noise level. Other considerations, such as whether the amplification is linear, then enter.

The great interest in the maser arises because it is a low-noise amplifier at microwave frequencies, and it is primarily for the amplification of very weak signals that it has been developed. In fact so low is its noise that background radiation noise is becoming important. In receiving a signal some sort of aerial system is used. Aerials can be made highly directional, particularly at microwave frequencies, but there is always the possibility that signals are arriving at the aerial from sources other than the transmitter (the transmitter itself will also be radiating a certain amount of thermal noise). For example, if the sky can be regarded as forming a black body enclosure at temperature T, the aerial would be receiving black body radiation from the sky, which would appear as noise. Actually the sky does not radiate as a black body and one of the uses to which masers are being put is to examine the distribution in direction and in microwave frequency of the noise arriving from the sky.

A further valuable feature of the maser is that although it has been developed primarily as a microwave amplifier there is nothing special about microwaves. There is little to be gained by trying to make longer wavelength maser amplifiers, as we already have low-noise amplifiers in this region, but there is nothing in principle to prevent our developing maser amplifiers for the near infra-red. There are limitations as the wavelength gets still shorter, for the maser will then no longer be a low-noise device, but it is an exciting prospect that before long we may be able to amplify in the infra-red. It can of course be objected that as we do not have infra-red oscillators there will be nothing to amplify, but this is not true, for any spectral emission line in the infra-red is equivalent to an oscillator. What we may not have is a tunable infra-red oscillator, but even here the maser may help, for it can be made to oscillate as well as amplify.

2. The basic physics

Our purpose in this section will be to review the fundamental ideas needed for understanding how a maser amplifies, and this will be done without specifying any particular frequency range.

Consider a system composed of a large number, N, of identical atoms placed in a large container, the walls of which are held at temperature T. Then after sufficient time has elapsed the atoms will be in thermal equilibrium with the black body radiation in the container. The relative number of atoms in any given energy level is then determined, by Boltzmann statistics, according to

$$\frac{N(E_1)}{N(E_2)}=\exp\left\{-(E_1-E_2)/\boldsymbol{k}T\right\}$$

where $N(E)$ is the number of atoms with energy E, each energy level being assumed to be non-degenerate. With $E_2>E_1$, $N(E_2)<N(E_1)$, showing that there are always more atoms in a lower energy level than in a higher level (in thermal equilibrium). We may now ask how this population distribution is maintained, for it is not a static distribution. A given atom may jump over a range of energy levels, but the transitions which are taking place must be such as to keep the population of any given level constant. The principle of detailed balancing states that in thermal equilibrium each transition scheme is itself in equilibrium. This is perhaps best illustrated by considering an atom which has only three energy levels (fig. 1) and a transition scheme which does not show detailed balance. Constant populations can be obtained if, in unit time, n atoms jump from E_1 to E_2, n from E_2 to E_3 and n from E_3 to E_1. Such a cyclic process, while maintaining constant populations, does not show detailed balance, for which it is necessary to have as many atoms going from E_2 to E_1 as from E_1 to E_2, etc. The principle can now be used to discuss what happens between pairs of levels, say the levels E_1 and E_2. Denoting the probability that an atom in E_1 makes a

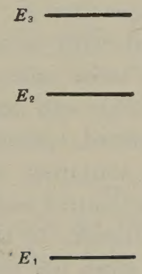


Fig. 1. Atomic system with three unevenly spaced energy levels.

transition to E_2 by p_u, and the probability for the inverse process, E_2 to E_1, by p_d, equilibrium requires that:

$$N(E_1)p_u = N(E_2)p_d$$

or

$$\frac{p_d}{p_u} = \frac{N(E_1)}{N(E_2)} = \exp\left\{-(E_1 - E_2)/kT\right\}$$

which is greater than unity. Thus the probability for a downward transition is greater than the probability of the inverse upward transition, to offset the population difference. The origin of the asymmetry can be found by supposing that the walls of the container are at absolute zero. Then an atom in E_1 is unable to draw energy from the black body radiation, as the radiation has no energy to give. On the other hand an atom in E_2 can fall to E_1, thereby releasing $E_2 - E_1$ to the radiation field (from which it will be absorbed by the container walls). This is the phenomenon of the spontaneous emission of radiation. At a finite temperature the black body radiation will be inducing upward and downward transitions with equal probabilities, say x, and the asymmetry arises solely because p_d is determined by x plus the spontaneous emission whereas p_u is determined only by x.

Suppose now that it has been possible to introduce into the enclosure a source of radiation which is capable of inducing transitions between E_1 and E_2, with probability S. The container and its contents will no longer be in thermal equilibrium and we cannot apply the principle of detailed balance. Let us suppose, then, that our atom only has the two energy levels E_1 and E_2. After sufficient time has elapsed a new equilibrium distribution will have been set up (not a thermal distribution) in which the number of atoms, $N(E_1)\{S+p_u\}$, which leave E_1 for E_2 in unit time is equal to the number, $N(E_2)\{S+p_d\}$, arriving at E_1 from E_2. Thus

$$\frac{N(E_2)}{N(E_1)} = \frac{S+p_u}{S+p_d}, \text{ with } p_u < p_d,$$

showing that $N(E_2)$ is again always less than $N(E_1)$, except that equality may be obtained when S is very much larger than p_d. A situation closely approximating to $N(E_1) = N(E_2)$ can be obtained with suitable transitions in the microwave region. The levels are then said to be saturated, for the power taken from the source is $S\{N(E_1) - N(E_2)\}(E_2 - E_1)$, and tends to zero as $N(E_1) \to N(E_2)$. It should be noted that the downward transitions induced by the source give energy to the source, not to the container walls. For any smaller value of S power is always absorbed from the source and delivered, ultimately, by emission stimulated by the black body radiation, to the container walls.

Next let us suppose that our atom has three unequally spaced levels, at E_1, E_2 and E_3, and that a powerful source of radiation is present to saturate the $E_1 \leftrightarrow E_3$ transition. We shall prove below that this can be done. Then $N(E_1) = N(E_3)$. There are now three possibilities for $N(E_2)$. It can be greater than, less than or equal to $N(E_1)$. If it is greater, and we consider the pair of levels E_1 and E_2, we see that there are more atoms in E_2 than in E_1. That is, as far as this pair of levels is concerned, a population reversal has been obtained. If $N(E_2)$ is less than $N(E_1)$ ($=N(E_3)$) a population reversal is obtained between E_3 and E_2. Now a pair of levels showing a population reversal has the very interesting property that on applying a signal to induce transitions between them the signal is not absorbed but is enhanced (amplified). This is simply because the induced emission, proportional to the population of the upper level, is greater than the induced absorption, which is proportional to the population of the lower level. Thus we have a way in which to get reversed populations and thereby achieve amplification. This is the basis of the maser: to obtain amplification by using stimulated emission from a reversed population.†

The above discussion of the 3-level maser is semi-quantitative, but it can be made much more quantitative as follows. We first introduce three probabilities, w_{12}, w_{23} and w_{13} which describe the $E_1 \to E_2$, $E_2 \to E_3$, and $E_1 \to E_3$ transitions under the black body radiation in the container. The $E_2 \to E_1$, $E_3 \to E_2$ and $E_3 \to E_1$ transition probabilities are then, respectively,

$$w_{21} = w_{12} \exp\{(E_2 - E_1)/\boldsymbol{k}T\}, \quad w_{32} = w_{23} \exp\{(E_3 - E_2)/\boldsymbol{k}T\}$$
$$w_{31} = w_{13} \exp\{(E_3 - E_1)/\boldsymbol{k}T\}$$

The saturating source (called the 'pump') induces transitions between E_1 and E_3 with probability P. The signal to be amplified, between levels E_1 and E_2, induces transitions with probability S between E_1 and E_2. Then, once equilibrium has been reached, the number of atoms leaving E_1, $N(E_1)\{w_{12}+w_{13}+P+S\}$, is equal to the number arriving at E_1 from E_2 and E_3, so that

$$N(E_1)\{w_{12}+w_{13}+P+S\} = N(E_2)\{w_{21}+S\} + N(E_3)\{w_{31}+P\}.$$

Similarly, considering the level E_3,

$$N(E_3)\{w_{31}+w_{32}+P\} = N(E_1)\{w_{13}+P\} + N(E_2)w_{23}.$$

There is no need to write down a similar equation for E_2, as this will be included when we specify the total number of atoms, N, as equal to $N(E_1)+N(E_2)+N(E_3)$. The solutions of these equations are algebraically unwieldy, but they simplify considerably under saturation conditions for E_1 and E_3. That is, when P is very large. Then

$$N(E_1) = N(E_3) = \frac{N}{2+\lambda}, \quad N(E_2) = \frac{N\lambda}{2+\lambda}$$

where

$$\lambda = \frac{w_{12}+w_{32}+S}{w_{21}+w_{23}+S}.$$

For reversal of populations between E_1 and E_2 we require that $N(E_2) > N(E_1)$, or $\lambda > 1$, or $w_{21}+w_{23} > w_{12}+w_{32}$. The condition can be written

$$w_{12}[\exp\{(E_2 - E_1)/\boldsymbol{k}T\} - 1] > w_{23}[\exp\{(E_3 - E_2)/\boldsymbol{k}T\} - 1]$$

which, for $E_3 - E_1$ large compared with $\boldsymbol{k}T$ (which is not necessarily the case in practice) becomes $w_{12}(E_2 - E_1) > w_{23}(E_3 - E_2)$. Some control over the magnitude of w_{12} and w_{23} is clearly desirable, and in this connection we must now mention an important extension of the theory. So far we have assumed that the atoms come into thermal equilibrium through their interactions with the black body radiation in the container. Usually, though, the atoms are contained in a solid, and are vibrating. The vibrations are excited by contact with the thermal bath and form what we may call a black body vibration field. It frequently happens that the atoms which are used in masers are more strongly coupled to this vibration field than they are to the radiation field, so that thermal equilibrium is primarily established via the vibrations. The only significant difference which this makes to the theory is to change the magnitudes of the transition probabilities. At present we do not know as much as we could wish to about the magnitudes, but there is the possibility that when we do we may be able to vary them as required.

---

† Maser is an acronym formed from the initial letters in Microwave Amplification by Stimulated Emission of Radiation.

The amplification is determined by

$$N(E_2)-N(E_1)=\frac{N(\lambda-1)}{2+\lambda}=N\left\{1-\frac{3}{2+\lambda}\right\}$$

so that to obtain large amplification we require λ to be as large as possible. But

$$\lambda=\frac{w_{12}+w_{23}+S}{w_{21}+w_{23}+S}=1+\frac{(w_{12}-w_{21})+(w_{23}-w_{32})}{w_{21}+w_{23}+S}$$

$$\simeq 1+\frac{1}{kT}\left\{w_{12}(E_1-E_2)+w_{23}(E_2-E_3)\right\}\Big/\left(w_{21}+w_{23}+S\right)$$

which suggests that it is advantageous to work at as low a temperature as possible. It should, however, be noted that w_{12}, w_{23}, etc. will also be temperature dependent. (At sufficiently low temperatures they probably will vary as T, but at higher temperatures they may show much more rapid variation.) It will be noticed that if S is large λ will be near unity and the amplification will be small. Masers may thus be expected to work best as low-power amplifiers.

It may be asked why the three levels need to be unequally spaced. If $E_3-E_2=E_2-E_1$ the pair E_2 and E_1 may provide amplification but the pair E_3 and E_2 will be attenuating and it will be correspondingly more difficult to obtain overall amplification. S may not be the same for E_3-E_2 as for E_2-E_1, and there are other complications.

3. General considerations

The above analysis for the 3-level case has been given in some detail because this type of maser seems to be potentially the most promising. There are other types, all depending on population reversals, but before going on to describe some of them it is of interest to consider certain additional features.

We have seen how to calculate the population reversal in the 3-level example, but an important question is how large a population reversal is required? Suppose then that we have an amplifying system in which the first stage is to be a maser amplifier, but in which we have not yet achieved a reversal in population. An incoming signal will be attenuated by the atoms in the maser material and also by the container in which the maser material is placed. The signal which is passed on to the next stage of the amplifier will thus be less than it would have been if the first stage had not been included. Now suppose that a population reversal has been achieved. The maser material is now emitting (negative attenuation) but there is still the container loss and no net amplification is obtained until the energy given out by the maser is greater than that absorbed by the so-called input losses. Having achieved net amplification, the excess energy must be dissipated in the output of the first stage (input of the second stage). It can however happen that the output stage is incapable of dissipating all the energy being released, in which case the stored energy builds up until some further limitation is reached. The maser is however no longer amplifying, but has gone into oscillation.

This process can be illustrated very clearly by means of a circuit analogue (fig. 2). L and C form a tuned circuit and V is an applied voltage at resonance. R_i represents the input losses and R_o the output losses. The maser is represented by a resistor R, which becomes negative as the maser becomes emissive. Prior to this it is positive. The power delivered to R_o is $R_oV^2/(R_o+R_i+R)^2$, and this must be compared with the power, V^2/R_o, that would be delivered to R_o if V were placed directly across it. There is no net amplification until R_i+R becomes negative, when the maser emission is greater than the input losses.

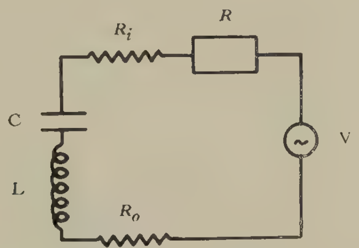


Fig. 2. Equivalent circuit of maser. R_i is input resistance, R_o is output resistance, and R represents maser resistance. In operation, R is negative.

However, if R is so negative that R_o+R_i+R is negative, the total circuit losses are negative, the stored energy builds up and the circuit goes into oscillation. Thus to obtain stable amplification the population reversal must lie between upper and lower limits.

A further feature can now be appreciated. To obtain large amplification R_o+R_i+R should be made as near zero as possible, while remaining positive. When, however, this has been achieved the circuit presented to the voltage source has very little net resistive loss, and has, therefore, a very high quality factor (Q) and a very sharp frequency response. Now in most applications one is interested in amplifying over a band of frequencies, which means that for a given amplifier bandwidth, Δf, the amplification must be limited. Taking

$$\frac{\Delta f}{f}=\frac{1}{Q}=\frac{R_o+R_i+R}{2\pi fL}$$

and the power amplification, G, as the ratio of the powers delivered to R_o with and without the maser stage,

$$G=\left\{\frac{R_oV^2}{(R_o+R_i+R)^2}\bigg/\frac{V^2}{R_o}\right\}=\frac{R_o^2}{(R_o+R_i+R)^2}$$

or

$$G^{\frac{1}{2}}\,.\,\Delta f=\frac{R_o}{2\pi L}=\text{a constant for the circuit.}$$

Thus to achieve high gain at a given bandwidth R_o needs to be as large as possible, and then $-R$ must be made correspondingly large. Inevitably some compromise is needed, and one of the main problems is to achieve high gain with an acceptable bandwidth; figures will be given later. One possibility is to spread the amplification over several stages. In terms of circuit analogues this is a straightforward problem, but using microwave techniques it is more difficult since it implies the amplification of travelling electromagnetic waves. This problem is currently being studied.

The amplification in a maser is obtained from a reversed population, and as this does not involve electron beams the sources of noise inherent with beams are absent. It was this particular feature which made the maser appear so promising as a low-noise amplifier.

A final aspect which deserves general comment is that in most masers the source of energy, usually called the pump, at frequency $(E_3 - E_1)/h$ in our 3-level example, is at a higher frequency than the signal to be amplified. As the maser is pushed to higher and higher frequencies this may prove to be a limitation as we may not have higher frequency generators. A number of schemes, usually of considerable complexity, have been suggested whereby lower frequency pumping may be used, but it is too early yet to know whether they will be successful. There is also the possibility that it will be possible to pump the maser with incoherent radiation obtained from a black body or other source. In the 3-level case it would be necessary to filter out the frequencies near $(E_2 - E_1)/h$ and $(E_3 - E_2)/h$ as these would tend to restore thermal equilibrium populations. Also, as the frequency is raised, the probability of spontaneous emission increases and this will show up in two ways. First it will be more difficult to obtain a reversed population and, having obtained it, the maser will generate more noise, for spontaneous emission at $(E_2 - E_1)/h$ appears as random noise from the maser material.

4. THE AMMONIA MASER

The first maser to operate successfully was built by a group led by Professor C. H. Townes, of Columbia University, New York. It used a beam of ammonia molecules in a most ingenious way to obtain a reversed population (fig. 3).

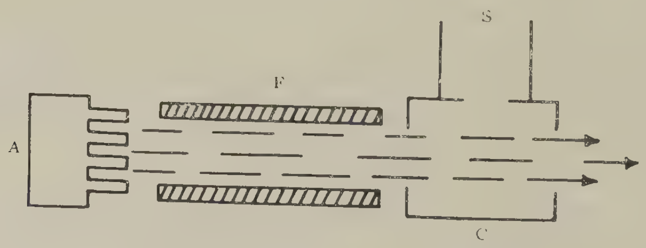


Fig. 3. Schematic diagram of the ammonia maser. A is the source of NH_3 molecules, F is the electrostatic focuser, C is the microwave cavity, and S is the signal input.

The ammonia molecule, NH_3, has the shape of a pyramid, the base being an equilateral triangle formed by the three hydrogen atoms, with the nitrogen at the apex. The structure is not rigid, and we can imagine that it is possible to plot a curve showing the potential energy of the molecule as the distance of the nitrogen atom, from the hydrogen plane, is varied. The curve will have two minima, one for the equilibrium distance above the plane and one for the same distance below the plane. If this was a classical system, and the nitrogen atom was placed in one of the minima, it would have to stay there. But being a quantum mechanical system the possibility of tunnelling through potential barriers exists, and the nitrogen atom will not stay on one side of the hydrogen plane, but will tunnel through the barrier so as to spend equal times on the two sides. One consequence is that instead of there being two equal energies, associated with the two equal minima, there are two slightly different energies. If we ignore excited vibrations and complications arising from centrifugal distortion due to rotation of the molecule, the two energy levels may be called E_1 and E_2, with the measured value of ν, in $\boldsymbol{h}\nu = E_2 - E_1$, being 23 870 Mc/s. $E_2 - E_1$ is a small energy compared with $\boldsymbol{k}T$, where T is any reasonable temperature, so that a beam of ammonia molecules may be expected to have almost equal numbers of molecules in the two energy levels, with a slight preponderance in E_1. It is known, however, that if the molecules are passed through an electric field they tend to be deflected in opposite directions. By means of suitably designed electrodes Townes was able to achieve partial separation: molecules in E_1 being deflected out of the beam while molecules in E_2 were concentrated in the beam. The separation, while not complete, was enough to ensure that in the emergent beam more molecules were in E_2 than were in E_1. Thus a reversed population was achieved. The beam was then passed through a resonant cavity, which is the microwave analogue of a tuned circuit, tuned to be resonant to $\nu = (E_2 - E_1)/\boldsymbol{h}$. The cavity was excited with radiation of frequency ν, to stimulate emission from the ammonia molecules, and amplification was obtained. It was found, however, that for a reasonable gain the bandwidth was extremely small, so that although the ammonia beam maser demonstrated that amplification could be obtained using reversed populations, it does not provide a very practical amplifier. If the beam intensity was increased sufficiently the maser went into oscillation (corresponding to $R_o + R_i + R < 0$), with a very stable frequency of oscillation. A good deal of work has since been done on the sensitivity of the oscillation to the variation of certain parameters, such as the beam velocity, the detuning of the cavity, etc. Some indication of the stability is obtained from the quoted results for two independent ammonia masers. For a period of one second the beat note was at 30 c.p.s. and was stable to 0·1 cycles, indicating a spectral purity of at least 4 parts in 10^{12}. Over one hour the maximum variation in beat frequency amounted to 5 cycles per second. The power output was of the order of 10^{-10} watts, which though small, is readily measurable.

5. THE 3-LEVEL SOLID STATE MASER

The 3-level solid state maser, of which we have already given the basic theory, was proposed in 1956 by Professor N. Bloembergen, of Harvard University (though similar ideas were also current in Russia). The proposal was rapidly taken up by a number of groups, and the first working maser was constructed at Bell Telephone Laboratories (1957). Since then a number of groups have reported successful operation.

The first practical difficulty was to find a suitable maser material. That is, a material having unequally spaced energy levels and favourable transition probabilities, for these determine the ease with which the 1–3 transition can be saturated and amplification at $(E_2 - E_1)/\boldsymbol{h}$ obtained. As a result of the investigation by microwave spectroscopy of many paramagnetic crystals, mainly carried out at Oxford by Professor B. Bleaney and his group, from about 1948 onwards, a number of substances were recognized as being possible maser materials. As progress has continued so new materials have been found and the most popular substance at present seems to be synthetic ruby, which is crystalline Al_2O_3 containing of the order of 0·1 per cent of the aluminium ions replaced by chromium ions. (The colour of many gem stones is due to impurities, chromium in ruby and emerald.) Apart from having the desirable properties of a high dielectric constant, negligible microwave absorption, and chemical stability, the Al_2O_3 plays no major part in the ruby maser. The active elements are the chromium ions. It has been found that unless the chromium ions are fairly well separated from one another interactions between them upset the maser action (the exact reason is still being investigated) and the role of the Al_2O_3 is to provide a matrix in which to carry the chromium ions.

In zero magnetic field each chromium ion is found to have two low-lying energy levels, each doubly degenerate, with a separation of 0·3826 cm^{-1}. In the presence of a magnetic field the degeneracies are removed and the ion has four energy levels, the spacings of which depend on both the magnitude of the field and its orientation relative to the crystal axes. An indication of the variation of the levels is given in figs. 4 and 5 for particular angles between the field and the c-axis of the crystal. Only three levels are usually used (the 3-level theory is easily extended to a 4-level case) and the ruby is oriented in a chosen field so that the spacings are as required. (The dependence on angle and field can be used to advantage in tuning the maser.) The ruby must be placed in a suitable enclosure so that it can be simultaneously exposed to the pumping

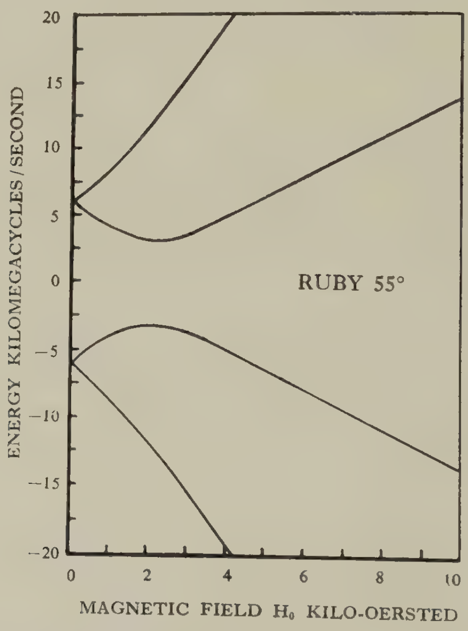


Fig. 4. Energy levels of Cr^{+++} in ruby in an applied magnetic field H_0 making an angle of 55° with the c-axis of the crystal.

radiation and to the signal to be amplified. If a resonant cavity is chosen it is usually required to resonate at both frequencies (fig. 6). As the radiation fields of cavity modes are polarized there are problems of exciting them and of ensuring that, at the position of the ruby, the field configurations are as required. There is nothing essentially asymmetrical about a maser, so that it is possible to interchange input and output. Difficulty arises if the output signal gets reflected back, for it then gets amplified again, appearing at the input, and so on. To avoid this happening it is usual to introduce asymmetry, by incorporating non-reciprocal elements, which allow a signal to pass freely in only one direction (the

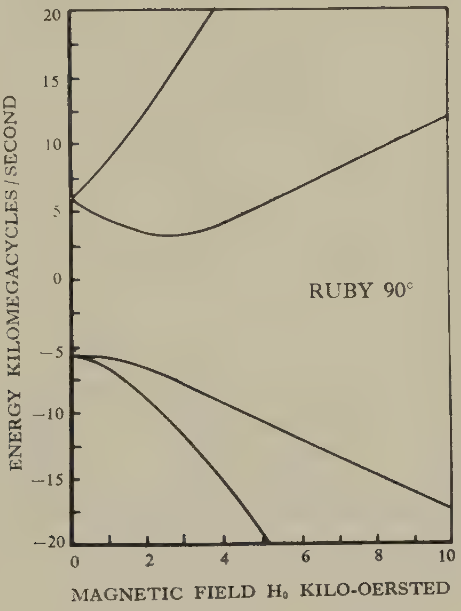


Fig. 5. Energy levels of Cr^{+++} in ruby in an applied magnetic field H_0 making an angle of 90° with the *c*-axis of the crystal.

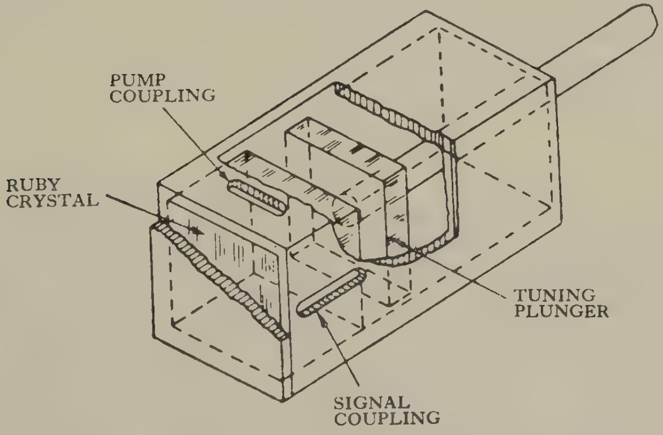


Fig. 6. Ruby maser of Morris, Kyhl, and Strandberg.

microwave analogues of one-way glass). In addition the cavity is usually immersed in liquid helium, which is cooled, by evaporation, to below its lambda point (2·17° K). The object is not only to achieve a lower operating temperature but also to take advantage of the altogether quieter behaviour of helium II. Above the lambda point evaporation of the helium causes ‘ bumping ’: below 2·17° K the ‘ bumping ’ is much reduced. It will thus be appreciated that having obtained one’s ruby there are still considerable technical problems to be overcome before one has a working maser amplifier.

An indication of the performance which can be obtained with a ruby maser comes from the work of Professor M. W. P. Strandberg’s group at M.I.T. They report the construction of an amplifier which is tunable over the range 8400–9700 Mc/sec with a *(gain)*$^{\frac{1}{2}}$ × *bandwidth* product of 40 Mc/sec, pumping being at 23 000 Mc/sec.

Other substances which have been used include lanthanum ethyl sulphate with gadolinium ethyl sulphate, and potassium cobalticyanide with potassium chromicyanide, where in each case the active material occurs last, the first substance being the matrix.

6. 2-level Masers

The problem of maser amplification depends on obtaining a reversed population and maintaining the reversal for as long as amplification is required. The ammonia and 3-level maser are good in this respect, for they will amplify indefinitely. For some purposes, as in radar, there is interest in amplifiers which work only intermittently and then for very short periods. In principle one might expect that, as a reversed population must correspond to a large departure from thermal equilibrium, it ought to be an easier proposition to obtain a reversed population for a short interval, than to retain it indefinitely. A number of ways are available for obtaining short-lived population reversals, and these are being actively investigated, particularly with 2-level systems. It is not easy to see what advantages they will have over successfully operating 3-level masers, and we mention them mainly for completeness. The substances being studied include phosphorus-doped silicon and various ionic crystals which have suffered radiation damage.

7. Conclusion

The work on masers which has been done over the past few years has definitely shown that low-noise amplification is possible using reversed populations. The initial experiments used heavy and expensive equipment, and much effort is currently being devoted to making a ‘ package ’ maser. That is, one which is reasonably cheap, light and robust and which does not require a team of highly trained physicists to run it. Commercial ammonia masers (oscillators) are already available. Unless the 3-level maser can be engineered successfully it is likely to remain a laboratory instrument, where of course it could be very useful. One of the more obvious difficulties is that laboratory masers usually operate at liquid helium temperatures. A very important technical break-through came when Mr. C. R. Ditchfield, of the Royal Radar Establishment at Malvern, succeeded, in 1958, in operating a ruby maser at 60° K, a temperature which can be reached by the evaporation of liquid oxygen. Since then even higher operating temperatures have been used. There is thus the prospect that this particular technical difficulty will be overcome.

An entirely different facet of the microwave amplifier situation is that, at about the same time that Bloembergen proposed the 3-level maser, a different type of microwave amplifier, the parametric (or mavar) amplifier, was proposed by Dr. H. Suhl, of the Bell Telephone Laboratories. This also promised to have low-noise features and, particularly, to work at room temperature. A good deal of effort has therefore been put into its development, with some success. Also, possibly because of the appearance of potential rival amplifiers, important progress has been made in the older travelling wave tube amplifiers. Thus the last few years have seen a great deal of activity in microwave amplification, and it is of interest to note that the new ideas have come, almost entirely, from the stimulus of academic research, carried out with little thought that sometime in the near future important technical uses would be found for apparently insignificant discoveries.

The developments in masers are so rapid that since this article was written ‘ package ’ travelling-wave-type maser amplifiers have been constructed and have been tested in the field for use in radio astronomy and communications using artificial satellites. Many of the technical problems have been overcome and there is little doubt that masers will prove to be more than laboratory instruments and will become commercially available. It also seems that they will be better low-power microwave amplifiers than their competitors, the parametric amplifiers, though they may be more expensive. The possibility of pumping at a frequency which is lower than the signal frequency has been demonstrated and it is understood than an infra-red maser has operated successfully.

REFERENCE

One book has now been published, a Methuen Monograph entitled *Masers*, by Gordon Troup. It gives a concise and comprehensive account of the whole maser field.

The Author:

Professor Stevens holds the Chair of Theoretical Physics at Nottingham University. He was educated at Magdalen College School, Oxford, and read mathematics and physics at Jesus College, Oxford. His interests are primarily in solid state physics, with particular emphasis on the study of magnetic phenomena.

O.E.E.C. Seminar on the Use of Television for the Teaching of Science

Under the direction of M. Henri Dieuzeide, who is head of Educational Television in France, the Organization for European Economic Cooperation held a five-day seminar at Ashridge College in July 1960. It was attended by educational authorities concerned with school science television broadcasting, heads of educational television programmes, science producers of schools programmes from 15 countries, and observers from the European Broadcasting Union, the Ford Foundation, the National Science Foundation (U.S.A.), UNESCO., the National Film and Television Council, and the International Scientific Film Association.

The Conference covered a wide range of subjects, from successful Italian teaching by the ' Telescuola ', which enables children to have lessons in areas where no secondary schools are available, to the use of television in secondary schools in Britain. Representatives from France showed examples from their educational programmes which go out to 5000 schools. Dr. John Baxter, of ' Continental Classroom ' (a series of 160 science programmes relayed at 6.30 a.m. from coast to coast in the U.S.A. for teachers of science) showed extracts from his repertoire.

Among the resolutions passed at the Conference were suggestions for— exchange visits by educational producers to train producers of newly established schools programmes; the exchange of information on science films, programmes and telerecordings; arrangements to promote co-operative production between European broadcasting organizations of schools science programmes; research into the use of closed circuits for the training of science teachers; and experiments in schools science programmes to be broadcast to whole areas of Europe where a common language is understood.

The International Scientific Radio Union

Dr. Reginald Smith-Rose, C.B.E., D.Sc., Ph.D., F.C.G.I., M.I.E.E., F.I.R.E., has been elected to succeed Dr. Lloyd Berkner of the U.S.A. as President of the International Scientific Radio Union (URSI). Previous holders of the office include the late Father Lejay of France and another British scientist, Sir Edward Appleton.

Dr. Smith-Rose, who has been Director of the Radio Research Station since 1948, retires from this post in a few weeks time. He recently succeeded Sir Lawrence Bragg as chairman of the Postmaster-General's frequency advisory committee and is a member of the newly appointed Committee on Broadcasting.

He was educated at Imperial College, London, and was a member of the staff of the National Physical Laboratory from 1919 and Superintendent of the radio division from 1939–47. He became Director of the Radio Research Station when it was established as a separate laboratory of D.S.I.R. in the following year.

The Neutron Booster

by M. J. POOLE

Atomic Energy Research Establishment, Harwell

SUMMARY

A sub-critical assembly of U^{235} has been used to increase by a factor of ten the neutron output of the pulsed source used for neutron time-of-flight measurements. This source consists of an electron linear accelerator producing a pulsed beam of high energy electrons which are then used to generate neutrons by X-ray production followed by photonuclear processes in a heavy element target. This paper describes some of the physics and engineering problems that had to be solved before a practical multiplying assembly could be built.

1. INTRODUCTION

The neutron booster is a multiplying target assembly which when bombarded by the electrons from a linear accelerator provides an intense pulsed neutron source for time-of-flight experiments, in which the energy of a neutron beam is found simply by timing it over a measured distance. Principally these experiments are concerned with the variation of nuclear cross-sections with incident neutron energy. The technique of using an electron linear accelerator to generate neutrons for time-of-flight work was pioneered at Harwell (Wiblin 1955, Goulding 1954) and to date three accelerators have been installed for this purpose. The earliest produced 120 mA of 3 MeV electrons during a 2 μsec pulse at a repetition rate of 400 p.p.s. Neutrons were produced by the reaction $Be^9(\gamma, n)Be^8$ ($Q = -1{\cdot}82$ MeV), the quantum radiation coming from the bombardment of a platinum target by electrons. The neutron output was 10^{12} n/sec during the pulse, giving a mean neutron output of about 10^9/sec. This accelerator was replaced in 1952 by a machine delivering 25 mA of 15 MeV electrons, again nominally in a 2 μsec pulse.†

Due to the higher energy it was now better to produce neutrons by photonuclear reactions in uranium, and the usual target for this machine was a block of canned uranium approximately 2 in. diameter by 3 in. long. This served the dual purpose of stopping the electrons to produce X-rays and absorbing these X-rays to give neutrons. The mean output from this accelerator was around 10^{11} neutrons/sec, while the peak output (assuming uniform neutron emission for ~ 1 μsec was 10^{14} n/sec. Substantial increase in neutron output may still be obtained by further increase in electron energy. The latest accelerator to be installed at Harwell has an electron energy of 30 MeV with a peak beam current of 300 mA, enabling a peak neutron output around $\times$ 5 10^{16} n/sec to be obtained from conventional targets.

By the use of a multiplying fissile target this output can be increased by a further factor of ten. The principle is simple: around the neutron target is placed sufficient fissile material to produce a fast neutron chain reaction but not enough to be critical. No neutrons are emitted until the pulse of electrons falls on the target, and then each neutron produced is multiplied in the fissile material, the exact amount of multiplication being determined by the quantity and kind of fuel used. In theory any desired degree of multiplication may be obtained by this process, but it will be seen below that in practice the factor has to be limited to about ten.

---

† Subsequent investigation showed that the *neutron* output was confined to about 0·7 μsec.

2. THE PHYSICS PROBLEM

In time-of-flight work neutron energy is obtained from the measurement of a time and a length. The energy resolution $\Delta E/E$ obtained is very closely given by

$$\frac{\Delta E}{E}=2\left(\frac{\Delta t}{t}+\frac{\Delta l}{l}\right)$$

where t is the time to cover distance l. The errors Δt and Δl are each the sums of errors at source and detector. Thus $\Delta t=\Delta t_D+\Delta t_S$; commonly, but not always, Δt_D and Δt_S are made equal by choosing pulse length and gate width to be equal, and frequently $\Delta l/l$ is negligible compared to $\Delta t/t$. Discussion will be confined to this case for the rest of this section.

At the same time the counting rate C in a given channel for the case of a detector which does not intersect the whole beam is given by

$$C\propto\frac{f\,S\,(\Delta t)^2}{4l^3}$$

where S is the peak source strength and f the pulse repetition frequency. Typical values for the quantities involved in the case of the neutron booster are $f=220$ p.p.s., $\Delta t_S=0{\cdot}25\ \mu$sec, $l=100$ metres, $S=10^{17}$ n/sec. Since, for a given resolution l is proportional to Δt we have the situation that $C\propto 2/\Delta t$. This means that to make the best use of a given source strength it is essential to keep Δt as small as possible and this consideration has proved a controlling factor in the design of the booster.

The basic physics problem then is to design a fast subcritical assembly which has the highest possible multiplication consistent with maintaining a short neutron pulse, and then to place around this sufficient moderator to bring a large proportion of the neutrons emitted into the energy range between 1 ev and 100 kev. This moderator must be arranged in such a way that it does not stretch the neutron pulse in time. Due to the finite life of a neutron diffusing in fissile material any multiplying assembly will lengthen the pulse to some extent. Using one-group diffusion theory it can easily be shown that if T is the characteristic relaxation time for fundamental mode neutrons, and τ is the mean life of a neutron in fissile material, for a bare fissile assembly

$$T=\frac{(M-1)}{(\eta-1)}\tau.$$

In this equation M is the multiplication defined as the number of neutrons *escaping* from the booster per neutron introduced, while η is the mean number of neutrons produced per neutron captured. The value of τ in pure U^{235} is 1·5 10^{-8} sec, while that in Pu^{239} is 0·84 10^{-8} sec. Unfortunately safety considerations ruled out the possibility of a plutonium booster on the Harwell site so that the longer value of τ associated with U^{235} had to be accepted. As a 0·2 sec electron pulse is used this immediately limits the value of M to around 10. The need to keep τ as low as possible also indicated that the average U^{235} density should be as high as possible—i.e. the amount of space taken up in the core by coolant and structural materials had to be kept to a minimum. It is also fairly obvious that moderated assemblies cannot be used. This follows from the fact that the absorption cross-section for neutrons in uranium does not increase inversely with velocity, and so a reduction in mean neutron velocity increases the mean life τ ($\tau = \Sigma_a v$, where v is the mean neutron velocity and Σ_a the mean macroscopic absorption cross-section).

A more refined calculation of T taking account of the presence of reflector and of the fact that the primary source is a point source rather than a distributed source shows that, of all the assemblies that can be made with a given value of M, the smallest T is given by a bare sphere of material with the source at its centre. As insufficient U^{235} was available to make a bare sphere, and as it was convenient to have some reflector present to adjust multiplication, the compromise was adopted of using a nearly spherical assembly of 37 kg 90 per cent enriched U^{235} surrounded by about 1 cm thickness of natural uranium.

A further source of time delays appears when we start to reduce the energy of the fission neutrons from the booster down to an experimentally useful value, which is accomplished by placing flat tanks, 2 cm thick, filled with water on three sides of the booster. Unless relatively slow (0 to about 10 kev energy) neutrons are prevented from returning to the booster core and there producing ' late ' fissions, then the neutron pulse will have a long ' tail ' which will spoil the resolution of any spectrometer. Thus cans 1 cm thick containing B^{10} metal powder had to be placed between the moderator cans and the core. This B^{10} allows fast neutrons to pass from the core into the moderator but absorbs any slow neutrons attempting to return to the core.

To summarize the eventual physics specification for the booster, we have:

(*a*) The booster must consist of pure fissile material of the maximum achievable density surrounded by the minimum amount of natural uranium reflector required to obtain the right multiplication.
(*b*) The source shall be at the centre of the booster.
(*c*) The multiplication must be limited to about 10.
(*d*) Moderator is required to bring the neutron output into the required energy range, but it must be separated from the core by a layer of B^{10}.
(*e*) Moderators must be excluded from the core.

A Monte Carlo calculation of pulse lengthening from all causes was carried out by Royston (Royston 1956).

3. THE ENGINEERING PROBLEM

Making a target to the above specifications presented an interesting problem to the engineers, mostly because of the necessity to keep the assembly small that is implicit in requirement (*e*) above. The electron beam carries a mean power of 5 kw, about 2/3 of which is dissipated in a volume of 1 cc at the centre of the booster. The remaining 1/3 is converted into X-rays which are then absorbed in the central part of the core. Finally, the fissions liberate a further 2·5 kw power distributed throughout the core volume. (Corresponding peak powers during the pulse are 26 MW for the electron beam and 13 MW for the fissions). The principal engineering problems then are

(*a*) How to remove 5 kw heat from the 1 cc volume of heavy element in which the electrons are stopped.

(*b*) How to cool the core without introducing light elements and without taking up more than 10 per cent of the core volume with coolant plus structural and canning materials. Furthermore, the coolant must have a slow corrosion rate on uranium so as to minimize the effects of can failure, must have a negligible corrosion rate for the can material and must not introduce unreasonable fire hazards.

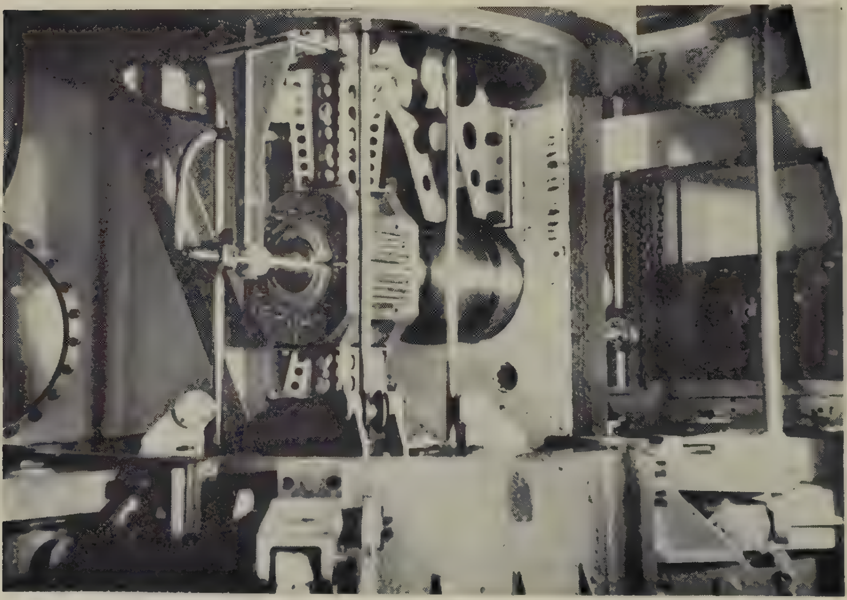


Fig. 1. Neutron booster during assembly, before fitting moderator tanks.

(*c*) How to assemble the core and reflector without any possibility of accidental criticality, and to ensure that even in the event of failure of any component accidental criticality during assembly and subsequent operation is impossible.

The solution finally chosen is illustrated in fig. 2. Here, C, D and E are stainless steel cans containing U^{235} made in the form of hollow cylinders and fitted coaxially one inside the other. These form the core which is cooled by mercury flowing through narrow annular gaps between the cans. Mercury was chosen as a coolant because it most nearly fulfills the conditions in (*b*) above. Extraordinary care was taken at all stages of manufacture to ensure that the 0·015 in. thick cans were flawless, and on completion each canned core component was thermally cycled under mercury at pressure before being accepted as sound. Through the centre of the core passes an assembly B which contains the target for stopping the electrons and then absorbing the X-rays produced to generate neutrons. Electrons from the left pass down an evacuated tube to the centre of the booster and then through a 0·007 in. thick stainless steel window into a stream of mercury. This mercury serves three purposes: it stops the electrons to generate X-rays, it removes the heat produced in this process and it cools the central U^{235} which absorbs the X-rays.

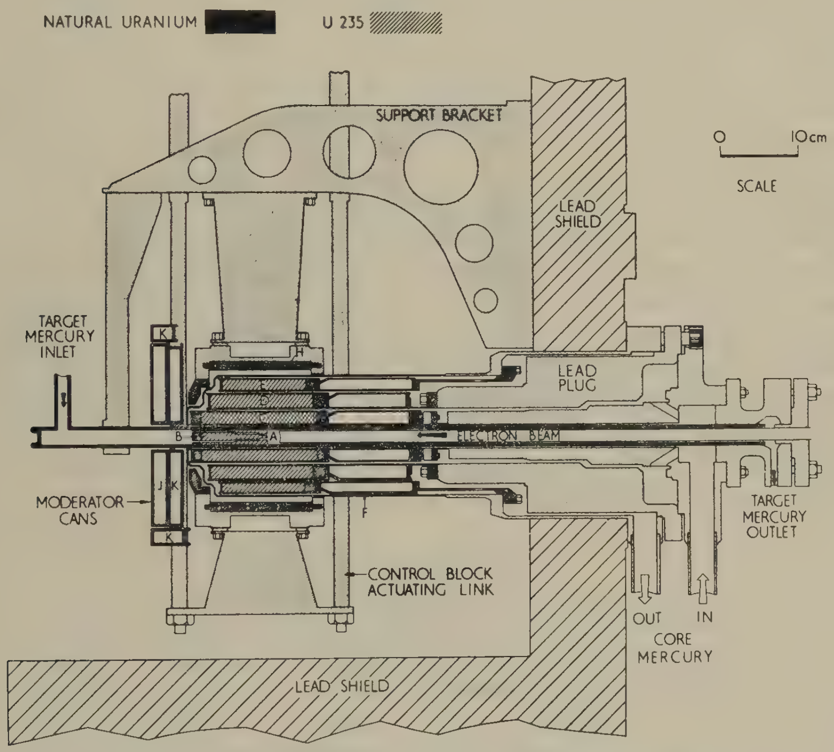


Fig. 2. Schematic diagram of the booster (not to scale).

Around this core and target is a reflector made of a shell of canned natural uranium rods H loaded into an upper and a lower carrier. The lower carrier can be moved away from the booster by a remotely operated control mechanism in order to adjust the neutron multiplication. This ‘ control block ’ can also be dropped freely under gravity in the case of any undue rise in neutron output and so acts as a safety mechanism during all operation and maintenance of the booster. Its principal use was during the initial assembly period when experiments were made to establish the exact effect on multiplication of the various external components and of the coolant. Outside the reflector are water tanks J with neutron shields K containing powdered B^{10} metal separating them from the core.

The booster is housed in a concrete cell 15 ft by 15 ft with 6 ft thick concrete walls, which cell forms part of the accelerator building. These walls are pierced by nine 2 ft diameter holes to provide neutron beams for experiments. When not in use these holes can be closed by flooding with water the space between their inner and outer aluminium windows.

4. MEASUREMENTS OF PULSE LENGTH FROM THE BOOSTER

An essential task to be undertaken on completion of the booster was an experimental verification that the all important pulse stretching effect was no greater than predicted. No instruments are available to do this directly so it was measured by observing its effect on the resolution of a time-of-flight spectrometer under conditions where the effect would be expected to be most pronounced. The 133 ev resonance of molybdenum Mo^{96} was observed in a transmission measurement using a 9·6 metre flight path, nominal 0·2 μsec primary neutron pulses and 0·1 μsec timing channels, and the observed transmission dip obtained using the booster as neutron source was compared with the corresponding dip when a non-multiplying target was used. These measured dips were then compared with the calculated dip for the resonance, making use of the known resonance parameters and the known characteristics of the spectrometer, and they were found to be compatible with the assumption of a rectangular electron pulse of $\frac{1}{4}$ μsec duration combined with a characteristic time for the booster of 0·09 μsec (Uttley, C. A. and Jones, R. H. 1960). Subsequently the electron pulse length has been measured with a fast oscilloscope and has been found to be $\frac{1}{4}$ μsec wide at half maximum height.

A full description of the booster has been presented to the second Geneva Conference on the Peaceful Uses of Atomic Energy (Poole 1958).

REFERENCES

GOULDING, F. S., HAMMERTON, J. C., KELLIHER, M. G., MERRISON, A. W. and WIBLIN, E. R. (1954), *Proc. IEE* **101** Part I, 248.

POOLE, M. J. and WIBLIN, E. R. (1958), Paper P/59; International Conference on the Peaceful Uses of Atomic Energy.

ROYSTON, R. (1956), A.E.R.E. Report No. T/R 2106.

WIBLIN, E. R. (1955), Proc. of first International Conference on the Peaceful Uses of Atomic Energy. Vol. 4, p. 35.

UTTLEY, C. A. and JONES, R. H. (1960), A.E.R.E. Report No. R/3281.

The Author:

Dr. Michael J. Poole was educated at Manchester Grammar School and Trinity College Cambridge, where he took his Ph.D degree. He is a Senior Principal Scientific Officer at A.E.R.E., Harwell. His chief experimental interests are in the measurement of neutron spectra.

The Sceptical Physicist

by R. A. R. TRICKER

Summary

Science teaching, both in school and university, is geared very closely to utilitarian aims. Has it also an important contribution to make to a liberal education? This article argues that it has, but that its opportunities are not being seized. Various ways of improving on the present state of affairs are discussed.

---

1. Learning how to think

The proper way to settle differences in a science is by reference to observation or experiment, and once settled it is not very easy to open such questions again. This fact is at the root of the difficulties which render most scientific questions so hopeless as material for discussion on a brains trust. Education, or at least the intellectual side of education, is largely a matter of learning how to think. How is this to be accomplished in a subject in which controversy is so difficult, in which a pupil can hardly hold an opinion of his own and be made to defend it successfully against the assaults of his teachers? For science to be a successful approach to an education, somehow this fundamental problem of the brains trust has to be solved.

We can learn accepted proofs and demonstrations as we learnt the theorems of geometry, in the hope that by repeating a sufficient number the general pattern will sink in and form a habit which will govern our thinking in the future. We all have to learn a good deal in this way, but if no more than this can be done it is but a poor thing compared to a subject in which the pupil can be plunged into the sea of controversy and taught to swim not only with but across or against the tide of opinion by developing views of his own. Is it possible in a science, and in particular in the physical sciences, to find situations which confront the pupil with the necessity for original thought—where he cannot rely upon or accept the opinions of others but must reach conclusions for himself? Even geometry would be a very indifferent discipline without its riders and problems. Riders and problems, however, do not furnish the complete answer to the question I am asking. They are essentially man-made and the fact that this is the case is sufficient to assure the enquirer that an answer does in fact exist. This certainty is absent in actual situations; its presence rules out an important possibility from consideration and its absence renders the situation more challenging.

Our much valued practical work too, essential as it is, rarely succeeds in meeting what is required; it fails to supply the essentials of a real situation. It consists of practical exercises rather than investigations. Too often the pupil has only to assemble the apparatus according to instructions, take the readings and insert the results into a formula provided for him. So long as the answer thus arrived at is within a certain range of the accepted value no further thought is usually asked of him. If the answer fails to fall within this range the measurements are sometimes repeated until the desired result is obtained. The apparatus is guaranteed to work by the teacher, and it is really the teacher who has performed the experiment. He has designed and modified the apparatus until it can be successful in the hands of the learner. He has given attention to those parts where care is necessary so that the required accuracy can be obtained, though occasionally success is obtained by a skilful balancing of errors of opposite sign. As a rule the more elaborate and costly the apparatus the less is the pupil required to think. If this scheme of things were to be carried to its logical conclusion laboratories might come to be furnished with a series of ticket machines so constructed that on the insertion of a sixpence in the slot the value of *g* or the wavelength of sodium light would be delivered to the student neatly printed on a card. The compilation and examination of records would be enormously simplified and the proceeds from the machines would help to solve the problem of the shortage of science masters.

Successful attempts have been made to remedy this state of affairs by occasionally presenting the pupil with an enquiry to be made which throws him much more on his own resources. He finds himself faced by situations in which he has first to plan his own procedure, to design, assemble and perhaps make some of the apparatus he requires himself. Then he has to test it and carry out any modifications which appear to be necessary and to use it for the purpose of the measurement or whatever it was that it was intended to do. Finally, there is no text book or book of tables to consult to see if the result is ‘correct’. The result has to be assessed in the light of the circumstances and its meaning discussed and elucidated.

There are many difficulties to be overcome in using this method. The first is to find problems which are of sufficient significance to engage the attention and energies of the student so that he is prepared to use his ingenuity to solve them. He is so obviously far removed from the frontiers of knowledge that to expect him to push them further out would be ridiculous. Nevertheless all of us are daily confronted by problems of a localized nature which we have to think about, which have never been investigated before and which are yet worth tackling. They may refer only to a particular machine or piece of apparatus such as a motor car, bicycle or camera lens and only to a particular aspect of their functioning. It is sometimes possible to arrange some of the usual measurements of the elementary laboratory in such a way that they can be approached with an open mind and require an original line of attack. If suitable problems could be found so that our pupils can be given this kind of experience at intervals during their educational career their education would be greatly benefited. It has probably to be admitted, however, that they are not so easy to find in the physical as in the biological sciences. In this lies an important argument against allowing the physical sciences to dominate the scene entirely as they now often seem to threaten to do.

The educational purpose of all learning, whether of facts, techniques or theories, is that what is learnt should form the basis for thinking. It is from the thinking more than from the learning that intellectual development proceeds. In science two different kinds of thinking are possible. The first we may call *creative thinking*, the second *critical thinking*.

2. CREATIVE THINKING AND CRITICAL THINKING

When we speak of training pupils to think in the field of science we usually mean the kind of creative thinking employed by scientists in their researches. It is the kind of thinking in which the niceties of strict logic are not always meticulously followed. In it the imagination plays a considerable part and logical jumps are by no means eschewed. It is the kind of thinking to which the investigations just discussed would lead in the main, and I would like to emphasize its importance rather than belittle it in any way. Of course hypotheses, views, or theories arrived at in this way have to be applied to other fields than those in which they originated—a process often thought of as testing. Those which survive are capable of becoming useful and are acceptable; those which prove incapable of application to other fields are replaced, if possible, by others which are, though where this is not possible they may be retained as a temporary measure until a better hypothesis can be found. It is natural that those people destined to become scientists should be exercised in this kind of thinking, but its value is by no means limited to them alone. Everybody in the course of everyday living is obliged to employ much of what thinking he is called upon to do in contexts where such scientific methods of thought are appropriate.

But it is important that the second kind of thinking which can be exercised in science and to which very much less attention seems nowadays to be paid, should not be neglected. Its educational potentialities are high—I would say higher than those of the first kind—and the second point to be raised in this article is designed to direct attention to them. To some extent it is the kind of thinking indulged in by the philosophers of science, though the problems they discuss are often very general and difficult and I would like to postpone the discussion of most of these until I raise my third point. Those which I have here in view include many very simple ones involving common principles of little difficulty.

The object of emphasizing the importance of critical thinking would be to generate a critical attitude in our pupils and to give them some powers of discrimination in scientific questions. Today too little distinction seems to be drawn between a systematic scientific study on the one hand and popular science on the other, and in the mind of the pupil the line between both of them and the pseudo-science of the more reprehensible advertisements and quack remedies is none too easy to draw. It is not my purpose to belittle the work of the writers of popular science. It is, in most cases, good in the sense that what is described is, in fact, correctly described. We all have to rely a great deal on such writings, but what we obtain is secondhand and rarely serves to sharpen powers of discrimination. Just as the object of a literary education is to furnish those who receive it with a measure of literary discrimination so those who receive a scientific education should obtain some powers of judgment and discrimination in scientific matters.

One of the greatest difficulties we face here is caused by the fact that by and large what we try to teach to our pupils are what we regard as the best of the scientific theories, those which are most thoroughly established and least open to question. In consequence it becomes very difficult indeed, for these immature people to come to any sort of assessment of what they are learning. Scientific theories, of course, vary very greatly among themselves. Some are very well founded and others extremely hypothetical. By concentrating our attention on the first group we deprive those we teach of the incentive which would lead them to be critical. It is becoming a characteristic of our times that scientists are set up as high priests, whose words are not open to doubt. Lack of the scientific knowledge which would enable people to judge is partly responsible, but the fault also lies partly in the dogmatic character of much of our teaching. Many people come to treat all scientific theories as of equal importance and equally worthy of acceptance and respect. My plea is for a more critical approach in science teaching; and if it is to be achieved it means that we must study something of which we can be critical. At present we are overwhelmed by the orthodox.

Attempts have been made in the past to achieve the ends here in view by studying some of the former theories of science, like those of phlogiston and caloric, tracing the reasons for their rise and decay. While these theories may attract the historian of science they are much less stimulating for most people than current views in a subject which is alive and controversial. I would suggest that a study of some of the theories of cosmology would be well rewarded. From the very nature of the case the evidence is limited and the theories inevitably incorporate a large hypothetical element. I would suggest that a comparison should be made, for example, between one of the well-established theories of physics, shall we say the theory of gravitation, supported by an enormous weight of evidence, with a theory like that of continuous creation, for which the evidence is very thin indeed. To help them in this process of assessment pupils might perhaps consider some of the arguments put forward by Professor Dingle in his presidential address to the Royal Astronomical Society.(1) It is equally obvious, of course, that evidence against the theory of continuous creation is just as difficult to come by and it is impossible for any definite conclusion to be reached. The virtues of a suspended judgment are not the easiest to emphasize in a scientific education. As a result of such studies I would hope that pupils might be brought to realise that scientific theories are neither born, nor do they grow up, equal. There is a hierarchy among them. The doubt which it is appropriate to attach to each theory varies. While no theory can escape it completely some give rise to it in large measure, and we do not have to accept the views of the moment of the specialist, only to change them with totalitarian unanimity when his fancy alters.

Logic is rarely included in the science course either at school or university. Whether or not it would be a good thing for it to form a separate branch of study to be taken as part of a scientific education is open to debate. This question, really subsidiary to my second point, might provoke a useful discussion. It could be argued, for example, that it would suffice for the logical questions at issue to be raised incidentally in the science course itself; in any case it would be necessary to do this to provide a backgronud to the fuller study, but if the latter is not attempted the incidental reference becomes the more important. It would hardly be disputed, I imagine, that the opportunities which the teaching of science presents for a critical study to be made of the nature of argument should be explioted. There is ample opportunity to do this, but the trouble about leaving it to incidental reference is that it is so easy to overlook it in the hurry to get on. I would like, however, to leave the question of whether it would be better to canalize this side of the work into a separate channel, an open one for discussion. There is much to be said on both sides.

Some might argue that it would be a mistake to run the risk of over-stressing the logical side of a scientific argument and they might emphasize the danger of getting bogged down in logic if we did. They might claim that so far attention to the nature of the arguments of physics has been sterile and led to little advance of knowledge. The latter stems from logical leaps which no logician could tolerate and science depends upon insight and imagination and not at all upon logic chopping and the niceties of argument. Even divested of their emotional content these arguments are, I believe, beside the point. They would be no more sensible than to blame art critics for not producing masterpieces of painting. The object of the critic is entirely different. It is to examine and evaluate rather than to fabricate; and even were his contribution to the advancement of science entirely negligible, which I am far from admitting, his contribution to education could still be immense. Furthermore, great insight and imagination are not possessed by many. Just as it would be foolish to hope to turn many people into masters in the art of painting, so it would be to expect them to become great thinkers in the field of science. But just as everybody can include in his education some understanding of art and acquire some ability, if only limited, to criticize, so no one can be considered to have received an education in science if no attempt has been made to develop some understanding of, and ability to criticize, scientific achievement.

Powers of criticism and judgment can be acquired partly by the cut and thrust between teacher and taught, whereby the latter is helped to set his own mind in order, partly by attention to the development of sound argument in the work being done at any moment and partly by the criticism of present theories as indicated above. My experience would suggest that the methods of many teachers are rather different from this. Their aim is not so much to develop judgment as to convince their pupils that there is a body of truth within their subject and to persuade them to accept it rather than to examine the situation critically and dispassionately. They are animated by an evangelical spirit, and this attitude tends to encourage acceptance and memorization rather than critical thought. It is in this way that the distinction between a study of science and popular science becomes blurred. In this, of course, teachers receive much encouragement from their pupils. Thinking is not a popular occupation; it is difficult enough to force oneself to think let alone encourage others to do the same. We all hope to acquire principles cheaply whereby our task of sorting things out may be made easy, rather than to face the hard work involved in reaching decisions for ourselves. However, if we are to acquire the habit of thinking for ourselves we must examine the arguments with which we are confronted. We must not allow ourselves to get away with unsound argument, though it is far from my intention to say that teachers should feel themselves personally responsible for the reasoning they are considering: rather it is the opposite. A dispassionate attitude rather than an evangelical one is essential. It is natural that teachers will impart to those who are entrusted to them their own enthusiasms and beliefs. My plea is merely that a passion for careful assessment should be among them. They must, by their example, lead their pupils to question and be on guard against faults of reasoning. That is why I think that the question of arranging for some study of elementary logic is worth raising.

It would be the object of any study of logic that the arguments of science should be assessed and be seen to pass muster. It would, I think, be possible to base a quite valuable study of logic based entirely upon arguments to be found in books on science. I had better emphasize that what I have in mind here is something essentially simple. Such vexed questions as the justification of induction which still perplex the philosophers merit a separate discussion which I will come to later. At the moment I am concerned solely with the development of acceptable arguments in the ordinary every-day study of science.

I once took down eleven elementary text books from my shelves and examined their treatment of electricity. What seem to me to be serious elementary faults were widespread.

The circularity entailed in the use of a moving coil voltmeter to demonstrate Ohm's law is very well known but the method was employed in one of the books. Four of the others cut the Gordian knot by simply stating the law dogmatically without attempting any demonstration at all. Two employed an uncalibrated ammeter to demonstrate Faraday's laws of electrolysis and a chapter or two later proceeded to calibrate the ammeter by electrolysis, using the very results it had been employed to establish. Eight of the eleven defined the volt in terms of work and used voltmeters freely throughout but gave no indication as to how they should be calibrated. This question of the calibration of a voltmeter seems to give rise to special difficulties nowadays. In the early stages of the study of electricity the only method of calibration open to us is a heating experiment assuming the value of the mechanical equivalent of heat from mechanics and heat. Until this calibration has been replaced by another any attempt to measure the mechanical equivalent of heat by an electrical method is bound to be circular. Yet four of the examining bodies include this specifically in their syllabuses for the Advanced Level of the G.C.E. in physics in spite of the fact that any other form of measuring potential difference in absolute units would be clearly outside the syllabus. It seems to be a common mistake to think that the heat measurement of Callendar and Barnes (their method is preferred) is all that is involved in the determination of the mechanical equivalent of heat electrically. The calibration of the voltmeter in absolute units, which is by far the most important and difficult part of the experiment, is overlooked.

Such examples could be multipled. Most elementary books on electricity introduce Ampère's law of action of current elements without experimental evidence in support. Mathematical treatises derive it from the theory of magnetic shells. But this theory is based upon the assumption that in its magnetic action a current may be replaced by a magnetic shell and for this in turn only the flimsiest of evidence is adduced. Current elements and magnetic shells are equivalent ways of looking at the problem and both require better evidence in their support than is normally brought forward in elementary treatments. What is involved here, of course, is the axiomatisation of science. The model is that of Euclid; its rewards and penalties are the same. We could end by making electricity a formal study divorced from reality, and the rise of various electricities would be as inevitable as that of non-Euclidean geometries. But let me return to some other difficulties before leaving these questions.

A form of circularity of argument which is becoming widespread is the use, in elementary stages, of concepts derived from later stages, the consideration of which depends upon an understanding of the earlier ones. Electric currents, for example are 'explained' as currents of electrons before any considerations have come to notice which might indicate a particulate structure for electricity. Atomic structure is invoked to 'explain' chemical reactions or electrical resistance. In illustration of what I have in mind I would like to give one or two quotations. The first is from the review of an elementary text book on electricity printed in the *School Science Review*.(2)

> "After reading the sentence on the front of the jacket, the reviewer opened the book with some interest and much misgiving. 'This is believed to be the first completely logical text book of electricity from the modern point of view . . . (The author) takes the view . . . that electrical studies may well begin at the proper place—the beginning. Accordingly he . . . starts with a lucid account of atoms, electrons and ions and the real nature, so far as now conjectured, of the electric current.'
>
> " The misgivings are well founded. 'Completely logical' means the deductive logic of Aristotle and Euclid; Galileo and Newton might never have lived (except that there would then be no electricity to write about). In fact, knowledge of atoms and electrons came at the end, not the beginning of a long chain of experiment and of inductive-deductive reasoning. (The author) turns electricity upside down and rewrites it as a deductive exercise. Chapter I is about electrons and protons. Chapter II, on Ohm's law, starts with electrons moving through the crystal lattice of a wire and defines the ampere as $6{\cdot}3 \times 10^{18}$ electrons per second. Chapter VIII starts by deducing from atomic theory what happens when currents are passed through electrolytes—and arrives at a theory of electrolysis which is many years out of date . . ."

A well-known and otherwise valuable book on chemistry(3) bases its theory on a large number of statements derived in the main from physics, for none of which is the slightest evidence adduced. Of the many that I counted in less than three pages of print the following are a few examples.

> *The electron.* The electron is a particle of mass $9{\cdot}11 \times 10^{-28}$ grms. This is only 1/1834 the mass of the hydrogen atom. The electron represents a charge of negative electricity. The electron *is* negative electricity . . .
>
> *The neutron.* The neutron is a particle of mass 1·0090 (on a scale on which the mass of the oxygen atom is 16) and diameter of the order of 10^{-13} cm. . . .
>
> *The proton.* The proton is a particle of mass 1·0081 ($0=16$). It carries a positive electrical charge equal in magnitude to the negative charge on the electron. Its diameter is of the order 10^{-13} cm.
>
> *The positive nucleus.* The positive nucleus of the hydrogen atom consists of a single proton. The positive nucleus of any other atom consists of protons and neutrons . . .

The same author's *Organic Chemistry*(4) is even more detailed. The following quotation comes from page six:

> " The electronic orbits of the hydrogen atom fall into types. A particular type of orbit in the hydrogen atom may be specified by assigning numerical values to each of two terms known as *quantum numbers*. The quantum numbers are:
>
> (*a*) *the principal quantum number n*, which may be any integer from 1 to 7,
>
> (*b*) *the angular quantum number l*, which for a given value of n may have any value from zero to $n-l$.
>
> The numbers of orbits of any particular type may be designated μ. μ is governed by the angular quantum number of the orbits and is given by the relation
>
> $$\mu=2l+1 \qquad \cdots\cdots$$

and so on.

If chemistry were to be studied only by those who had already graduated in physics this approach would be not so bad, but the book is written for elementary students whose knowledge of physics is likely to be even more sketchy than their knowledge of chemistry. To them it cannot be anything other than mumbo-jumbo. Also elementary chemistry contributes an important quota to the basis for these theories of the atom and circularity of argument is clearly not avoided. It is a pity that books otherwise so valuable should be marred in this way.

The use of such 'advanced' concepts is sometimes defended on the grounds that the evidence supporting them is to be found in the phenomena they are capable of explaining and that, as we cannot study them all at once, it does not very much matter where a beginning is made. This is a half truth. It ignores the fact that science does, in fact, possess some logical structure. One cannot properly discuss, for example, the energy states of the atom before considering the evidence for the electron and the other particles, and in turn the electron before studying the more elementary and general properties of electricity. Furthermore the method gives no insight into the mind of the originator of the concepts and the problems he faced. The picture presented by treating modern theories as *faits accomplis*, without tracing their growth, is too implausible as an explanation of isolated facts. It is all very well to say let us make these assumptions and verify them *a posteriori* but we all know that they are so numerous and so interwoven that the attempt is never made and we end up with our minds in a muddle knowing neither what we believe nor the reasons we have for believing it. This, of course, is also the answer to those who say that our pupils will inevitably have heard of atoms, electrons, neutrons and the rest and why should we not make use of their knowledge? The mistake is to confuse the results of hearsay with adequately investigated knowledge. We all acquire information through hearsay, but we can only order our thinking properly if we draw a clear distinction between it and those things we have studied and can reckon to some extent to judge.

It would not be difficult to illustrate other more or less elementary faults of logic from the kind of arguments used in elementary science. The search for implicit assumptions would be rewarding not only as a mental exercise but it might well yield valuable criticism and throw light on the development of the subject as well. When, for example, Dalton said, " For thou knowest that no man can split an atom " he was begging a very large number of questions. You cannot alter the world of nature by making definitions. The use of vast extrapolations inevitably involves assumptions of uniformity. To take the red shift in the light from the distant nebulae as evidence of an expanding universe is obviously to assume that in the course of travelling for a thousand million years from one of the distant nebulae a train of light waves does not alter its frequency, that atoms a thousand million light years away vibrate to the same frequencies as here in the laboratory and that they did the same a thousand million years ago as they do now. The universe cannot be measured with a measuring rod. We are at the mercy of large extrapolations concerning the properties of light and there are fundamental difficulties over the measurement of time. A full discussion would go far beyond the elementary sphere but the necessity to be on guard is clear enough.

3. The case for the study of philosophy

So far we have been concerned with questions of tactics, with the arguments we employ in detail. All I have been asking for is a well-reasoned course (the need for which in principle probably few will wish to question) and also that we should try to cultivate clear and critical minds. But questions of tactics merge into those of strategy. Inevitably we would at some point be brought into contact with the problems of induction, of existence and of perception. The futility of attempting to provide a deductive basis for induction is recognized and the approach through probability, though promising, is still beset by difficulties. We still require an additional principle besides those we find in formal logic and this brings me to what I would like to make my third and last point. It is this: granting that we agree to encourage pupils to think and we devise situations for them in which they can successfully hold views of their own, and that both in this and in approaching the established theories of science they should look critically at their reasoning, would it not still be a valuable study to introduce them to the problems of the theory of knowledge? Interest in meaning is tending to be reduced. " It is found that the physicist or engineer can do all that he needs when working with particle accelerators if he has familiarized himself with the relativity equations and the rather specialized notation they employ. Pre-occupation with meaning, such as was common in Eddington's day, is no longer necessary. It suffices for practical purposes to know how to use the letter symbols."(5) Should we not raise such questions of meaning? What is the nature of the concepts used in science? What meaning, if any, is to be attached to the question " Do magnetic fields exist? " What are the difficulties of the position of naive realism occupied by most scientists? What of the problems of solipsism which have to be faced by empiricist and idealist?

It seems to me that it would be a great pity if these questions ceased to be asked. " The criteria of a man's education are not only the questions he can answer but even more those he asks."(6) One would be prompted to demand what sort of an education it is which science provides if these and similar questions were ignored. Can it indeed be said to provide an education at all if it does not make contact with its philosophy? Is not science too closely concerned with the elucidation of facts and the ordering of them into axiomatic systems and neglectful of the purposes of so doing? Are indeed the purposes of the study of science purely utilitarian? To admit that they are would undermine any claim the subject might have to be considered a vehicle for any education worth the name. It is difficult to overlook the fact that recent debates have failed to stress any argument other than the utilitarian for the widespread increase in the study of science which is desired.

One of the important ingredients in the greatness of the Oxford ' Greats ' course is its content of philosophy. Is there not a place for a ' Science Greats ' in the sense of a combination of science and philosophy? This would be something different, it must be noted, from a mixed degree including subjects drawn from both the arts and sciences. I do not wish to discuss the merits or limitations of the latter idea. Properly conceived, courses in science and philosophy would have much to contribute to each other and the study would, in a very fundamental sense, form a unity. It could, moreover, make an important contribution to closing that gap which at present separates those engaged upon the sciences on the one hand and the arts on the other.

A study of philosophy would, however, be of particular value to the scientists. It could introduce them to questions which can be discussed but to which no definite answers can be given. The scientist is too often prone to dismiss these as being meaningless and yet they include some of the most interesting that have intrigued the mind of man. Some of them are very relevant to the pursuit of science itself. The natural entry into philosophy for a scientist would be through epistemology and a consideration of scientific method. An early insistence upon simple questions of logic would prepare pupils for such considerations, but it would not be possible to start detailed study of philosophy at a very early age. Whether school or university should be responsible for beginning it is a question which requires discussion and this can only be based upon experience gathered through experiments. Some made towards the later stages in the Sixth Form would be very valuable here.

While scientists could enter upon some study of philosophy using their scientific studies as a basis, the reverse might well be the case for those who, for want of a better name, are called non-scientists. At early stages—in the Middle School—they should, of course, have made a study of science of a more or less empirical nature and become familiar with a range of the phenomena with which science deals, but as they get older this approach loses its appeal—they have elected to concentrate their efforts in other directions. Catalogues and accounts of technical achievements or the latest theories strike no chord since the basic understanding which would enable them to judge of their value, is lacking. An approach to science through philosophy might be much more stimulating and profitable. What could be done at school and what at the university could again only be settled by the results of experiments, but I am far from being convinced at present that a start on simple lines at the later stages of the Sixth Form is out of the question. The problem is important enough to warrant it being given consideration. If we are to achieve any measure of unity between the disciplines which at present divide educated people it is surely to be found in their philosophy which is all that they possess to any extent in common.

This article has been written to direct attention to what I think are the most important questions in connection with the use of science as a vehicle for education. Many share the feeling that all is not well with our teaching of it. If we are not to lose many of those qualities which have enabled us so far to hold our own in a world in which we are not only heavily outnumbered but in point of material assets out-gunned as well, these questions deserve urgent consideration.

REFERENCES

(1) DINGLE, H., 1953, *Monthly Notices of the Royal Astronomical Society*, **113**, 3.
(2) *School Science Review*, 1949, **30**, 111.
(3) DURRANT, P. J., *General and Inorganic Chemistry*, Longmans.
(4) DURRANT, P. J., *Organic Chemistry*, Longmans.
(5) KAPP, R. O., 1959, *British Journal for the Philosophy of Science*, **10**, 37, p. 3.
(6) BICKLEY, W. G., 1958, *Mathematical Gazette*, **42**, 340.

The Author:

Dr. R. A. R. Tricker, H.M.I., is a Staff Inspector in the Ministry of Education.

The Tandem Generator

by W. D. ALLEN
Atomic Energy Research Establishment, Harwell

SUMMARY

A brief description is given of the development of the electrostatic generator from the single-ended machine to the double-ended, or tandem generator. The performance of the tandem machines is illustrated with reference to the U.K.A.E.A. tandems.

The tandem generator is the latest development of the electrostatic generator, which is often associated with the name of its inventor, Robert J. van de Graaff. A van de Graaff machine is rather like a Wimshurst machine, with a rubber belt taking the place of the glass disc. A schematic of a machine working by self-excitation is shown in fig. 1: charge is sprayed on the belt at earth potential and carried up into the top cap, where it is taken off a second set of spray combs to the cap. The cap is a Faraday cage, inside which electric fields, except those on insulators, cannot exist: therefore, although the voltage of the cap will rise steadily, owing to transfer of charge to it, the operation of the top spray comb will be unaffected. Demonstration models of the type shown in fig. 1 are easily made and are available commercially. They will generate sparks of several inches in air, corresponding to top terminal voltages between 100 and 300 kv.

When used as a particle accelerator, of course, the machine needs to be modified in various fundamental ways from the machine of fig. 1. Everywhere the electric fields must be graded as smoothly as possible: this means that not only must the top terminal be suitably shaped, but down the belt the field must be graded by a set of plates whose potential is determined by a resistor chain. Fig. 2 shows a picture of the Harwell electrostatic generator, with the pressure vessel removed. The pressure vessel enables the machine to be operated in high pressure gas : this not only greatly increases the electrostatic strength of the gas: but enables one to keep the gas very dry—a point which anyone attempting to operate a model of the type of fig. 1 will appreciate. As a particle accelerator, the top terminal is run positive, and positive particles, e.g. protons, are accelerated from an ion source down a highly evacuated tube to the earthed end. Here the beam is deflected through 90° by a magnet, and passes between a pair of slits, which are differentially connected to an amplifier. The output of the amplifier goes back to some electrically controllable load on the terminal, e.g. through a set of needles facing the terminal. If the voltage of the machine rises, the beam will hit one slit more than the other, and the amplifier sends a signal of sign such as to increase the load on the machine. In this way, voltage stabilities of one kilovolt in several million volts can readily be achieved.

With this introduction, we can understand the virtues and limitation of the electrostatic generator. Effectively the only other machine delivering positive particles in the 1 to 5 million volt range is the cyclotron. Now a 5 MEV cyclotron will deliver more current and will be cheaper than a 5 MEV van de Graaff. Despite this, the van de Graaff is a more useful machine, because as compared with the cyclotron it has precision and flexibility. Of its precision we have already spoken: its flexibility arises from the fact that the voltage can readily be varied by varying the injected spray current. It is also flexible in that any one of a variety of gases—hydrogen, helium, deuterium—can be introduced into the source. However, there is a major limitation. Whereas cyclotrons can be built to deliver protons of energy up to 750 MEV, electrostatic generators can reach at most 10 million volts without a tube, while, with an accelerating tube, voltage is limited to about 6 million volts. Two large installations have reached slightly more, but in general the average installation, with stack length 12ft and pressure vessel diameter 8ft, will deliver particles with a maximum energy of 6 MEV.

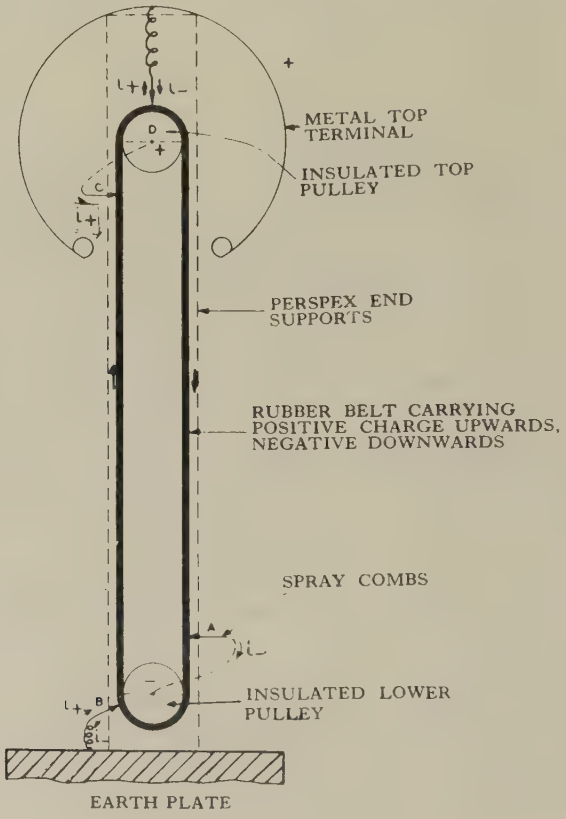


Fig. 1. Section of machine illustrating electrostatic generator principle.

The single-ended generator of the type we are discussing was first described by van de Graaff in 1931, and he and the early experimenters confidently looked forward to reaching 10 million volts with the machine. Now, there is a good reason for making the range 15–20 MeV a final objective. Coulomb barriers for protons are of the order of 1 million electron-volts for the nuclei of light elements (e.g. lithium) and 15 MeV for those of heavy elements (e.g. uranium). A cursory survey of what is known in detail about nuclei –energy levels, spins, parities, etc.—would show a vast amount of information at the light end of the periodic table and only sparse information about the middle and heavy end of the table.

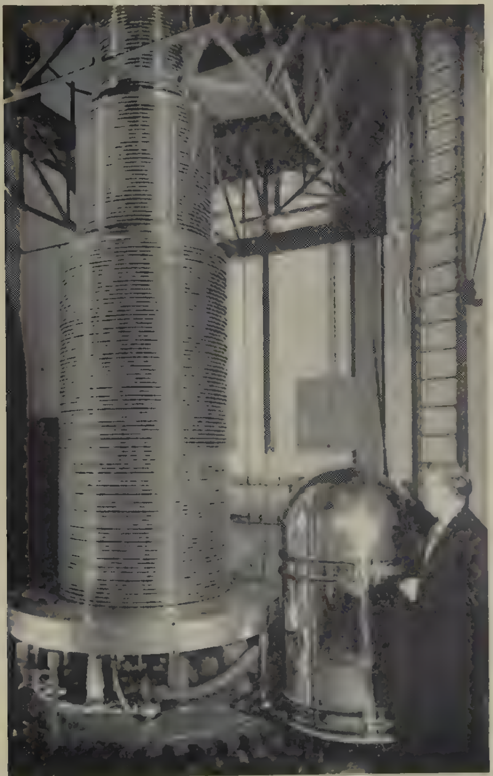


Fig. 2. Photograph of Harwell single-ended van de Graaff machine with pressure tank removed.

The reason for this is simple. Elements of mass up to about 56 (iron) can be studied by proton beams from electrostatic generators (and Cockcroft-Walton sets, for the lightest elements), since up to iron the beam has sufficient energy to surmount the Coulomb barrier. For the heavy elements, results are few, and are derived mainly from radioactive decay studies, Coulomb excitation and neutron inelastic scattering. It is clear that electrostatic generators working in the MeV region between 5 and 20 MeV, would operate in a very profitable field.

The goal of 10 MeV is now comfortably to hand, but not by way of the single-ended machine. In 1937, two far-sighted individuals, W. H. Bennett and H. Kallmann, took out patents on a new technique in particle acceleration which depended on doubling the voltage through which the particles were accelerated. The principle they described depends on the properties of the hydrogen atom, which in its neutral state consists of a positively charged central nucleus (the proton) and a negatively charged planetary electron. Most accelerators of heavy particles depend on the removal of the electron, which is the usual ionization, and the acceleration of the residual positive ion (the proton in the case of hydrogen). However, it so happens that the hydrogen atom can not only lose an electron, but can at a suitable velocity pick up an additional electron from another stationary neutral atom, thus forming a negative hydrogen ion composed of a proton and two electrons. These negative ions can be attracted to a central high voltage positive electrode, where they are passed through a small quantity of material—a thin film of light material, or a tube containing gas. At their higher velocity (a result of the acceleration) the negative hydrogen ions tend to lose their electrons, leaving only positive protons. These positively charged particles are then repelled from the positive electrode, and arrive back at earth potential with an energy corresponding to accelerating a single charge through twice the voltage on the central electrode.

The idea lay fallow for many years, although it was independently revived by L. W. Alvarez in 1951. It was not taken seriously, perhaps, because no one believed that negative ion sources of sufficient intensity could be produced. However, in 1955 it was announced from the laboratory in Wisconsin of R. G. Herb, one of the foremost pioneers of electrostatic generator techniques, that negative currents of 25–35 μA had been obtained. This immediately meant that the double-ended electrostatic generator, became a practical possibility. The ‘ Swindletron ’ as it had been light-heartedly called (‘ two volts for the price of one ’) now became a respectable member of accelerator society, and was renamed the ‘ tandem generator ’ (two electrostatic generators in tandem). In America, the Chalk River laboratory placed a contract with the High Voltage Engineering Corporation for a horizontal machine. This machine was first tested in April, 1958, and made a spectacular debut, yielding a current of 0·3 μA of protons at 10 million volts soon after the first switch-on. At about the time that the Chalk River contract was signed, i.e. early in 1956, interest was growing in the United Kingdom in the possibilities of tandem generators. In the United Kingdom Atomic Energy Authority, this interest arose, not only because of the very wide possibilities of these machines in basic research, but because they can be used to study the level structure of the transuranic nuclei and as potential sources of neutrons in the 8–12 MeV range. Therefore, in June, 1956, an agreement was reached whereby there would be two machines, one at the Atomic Weapons Establishment, Aldermaston, and the other at the Atomic Energy Establishment, Harwell. The machine parts were to be manufactured by the Metropolitan-Vickers Electrical Co. Ltd., of Manchester, the ion sources and strippers by A.W.R.E., and the accelerating tubes by A.E.R.E. The joint project proceeded smoothly, and both machines delivered their first beams at Easter, 1959. There are now many tandems built or building in the world, but it will suffice if we illustrate the machine in detail with reference to the A.E.A. machines.

One fundamental difference between the United Kingdom and American machines is that the U.K. machines are vertical. This was dictated in the first place by the fact that horizontal machines have their own particular engineering problems, of which there was no previous experience in the United Kingdom. However, the vertical generator has other advantages. One is that weight and size in the centre terminal are not limited, in the way that they are in a horizontal machine. There is, however, another feature of the vertical machine which is put to full use at Harwell (fig. 3). For a vertical machine, deflection through 90°, which as we have seen is essential to the process of stabilization, brings the beam into the horizontal. The magnet can then be rotated around a vertical axis, and deliver the beam in turn to a series of target rooms : at Harwell three target rooms are disposed in a semicircle, as shown in fig. 3. This means that, while an experiment is in progress in one target room, experimental equipment can be set up or altered in adjoining target rooms.

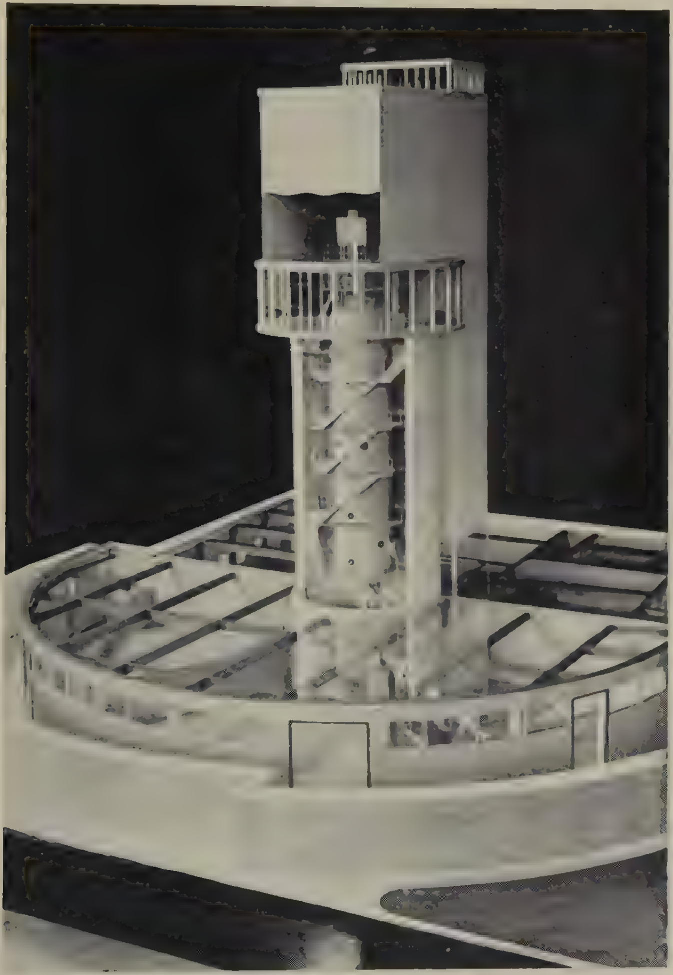


Fig. 3. Photograph of model of Harwell tandem generator and building.

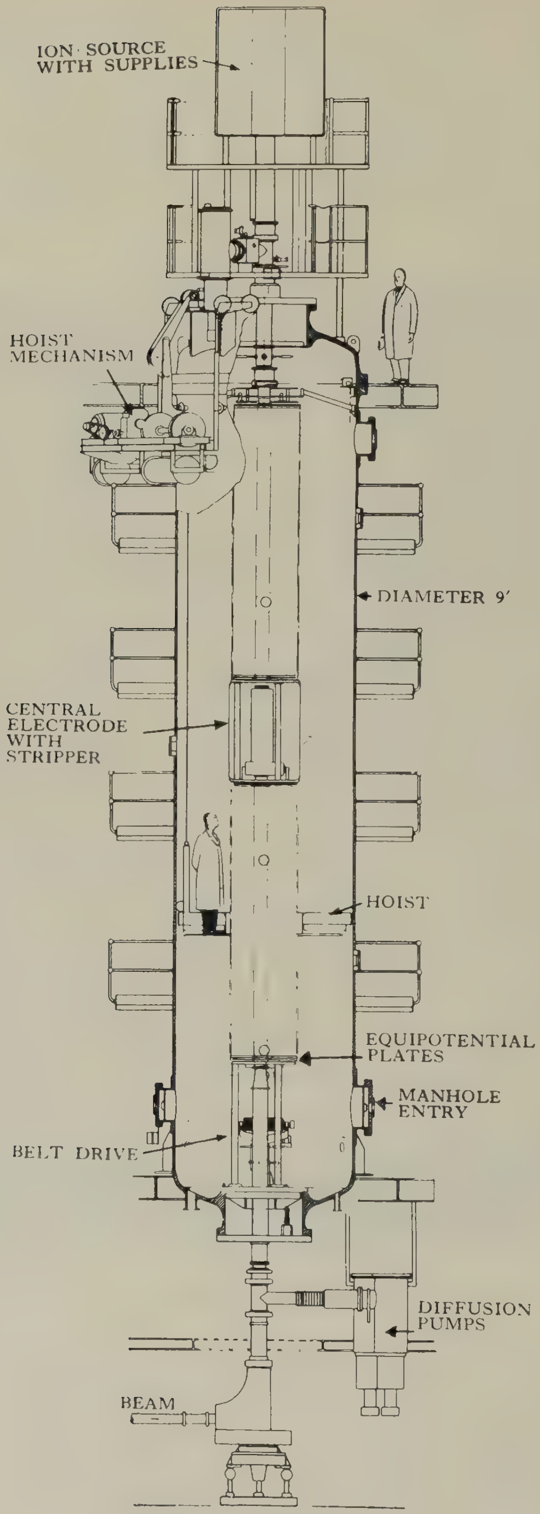


Fig. 4. Vertical section of A.E.A. tandem generator.

Fig. 4 shows an outline drawing of the United Kingdom machines. The electrostatic generator itself is housed in a steel vessel 50 ft long and 9 ft in diameter weighing 45 tons. When the vessel is fully pressurized, the weight of gas is 2 tons. Access to the machine in general is via manholes, usually from below : access to the stack is by way of a hoist, whose cables and ropes are lowered through portholes at the top. The belt drive is simple : it is an inverted motor, in which power is delivered to the central stator, which drives the rotor external to it, the rotor being also the belt pulley. The stack of equipotential plates and the accelerating tube are built to close tolerances : for example each tube, 14 ft in length, of 154 electrodes and glass rings, is made so that when erected on a horizontal base, no part of any electrode rim deviates by more than ± 0·020 in from the related mean vertical position.

The ion source is very simple : it is a radiofrequency type positive source, with the collecting voltages external to the source reversed. Charge exchange takes place in the exit canal of the source, so that positive ions, neutrals (fast and slow), negative ions and electrons all emerge. A positive collecting field rejects the positive ions, while a magnet deflects the negative ions through a small angle. Neutrals are removed since they are undeflected, while electrons are removed by very strong deflection. Beam currents into the machine of up to 10 μA are normal, although 50 μA have been attained at A.W.R.E. The stripper at A.W.R.E. is a tube carrying gas, $\frac{1}{8}$in diameter and 2 ft long, i.e. of the same type as the American stripper. The transmission through such a tube of some 50% of the beam represents considerable aligning accuracy. At A.E.R.E. the stripper is a thin foil of carbon, 1 cm in diameter and 4 micrograms/cm^2 in thickness. This thickness is only slightly thicker than optimum for stripping ; the foils are almost completely transparent. Foil life has not been a problem (36 foils can be mounted on a wheel at one time), and the use of foils greatly simplifies alignment and other problems.

The machines have been operating up to 12 MeV, although at Harwell difficulties with the belt have been a hindrance. Currents on target are normally a few microamperes : it is to be noted that, under optimum conditions, it is possible to inject say 8 μA into the top of a machine, accelerate the beam through 60 ft vertically, deflect and transport it 40 ft horizontally and deliver 3–4 μA on target. Another interesting variant is the use of oxygen. When O^- (negative oxygen ion) is injected into the machine, it is accelerated to say 5 MeV. Stripping will remove several atomic electrons, leaving say O^{+++++}, which on acceleration through 5 M.V. will give an ion of total energy 30 MeV. The field of Coulomb excitation using oxygen ions looks very promising.

Although the machines have been operating for only a short time, several new fields of interest are rapidly opening up. Limitation of space precludes detailed description, and one example must suffice. This is in the field of photodisintegration, where a nucleus absorbs a γ-ray quantum and emits a proton (γ, p), or a neutron (γ, n). In the period 1950–1955, workers with betatrons and synchrotrons in the range 10 MeV to 30 MeV showed, in their studies of photodisintegration, that there was in all elements a ‘ giant resonance ’: the cross section for (γ, p) or (γ, n) rose to a high peak at about 20 MeV, with a width of about 6 MeV. Difficulties of technique prevented accurate studies of these resonances. With the tandem generator, one can begin to study the inverse reaction (p, γ) e.g. one bombards a target with protons of energy up to 10–12 MeV and looks for a gamma ray of energy of about 20 MeV. The much greater resolution of the tandem generator beam makes it possible to look for a fine structure in the peak. Studies to date show that

$$B_{11}\ (p, \gamma)\ C_{12}$$

shows a smooth peak, while

$$N_{15}\ (p, \gamma)\ O_{16}$$

and

$$Al_{27}\ (p, \gamma)\ Si_{28}$$

show very marked structure effects.

In all, the tandem generator is a machine which is of interest in itself, and which will certainly make a major contribution to nuclear physics in the 5–12 MeV range.

The Author :

Dr. W. D. Allen was educated at Adelaide, S. Australia, and Oxford. His fields of interest have been waveguide research in radar; electromagnetic separation, at Berkeley and at Harwell; and the use and development of electrostatic generators, including the measurement of fast neutron fission cross sections and the construction of the tandem generator.

The Physical Science Study Committee
(3) The Planning and Structure of the Course

by STEPHEN WHITE
The PSSC, Cambridge, Massachusetts

1. The origin of PSSC

The Physical Science Study Committee was organized in response to a widespread dissatisfaction among physicists and physics teachers with the state of secondary-school physics in the United States. The course devised by the committee will enter into general use during the academic year 1960–61. This paper will attempt to set forth the content of the course and to account for the form the course has taken.† As a preliminary, however, it seems advisable to discuss briefly the place of physics in the American secondary education scheme, and to offer an account of the sources of dissatisfaction.

The American student is given nominal instruction in science from his earliest years, but his first acquaintance with any one of the sciences as a coherent discipline comes during his years in grades nine to twelve (secondary school). During this portion of his schooling, he is likely to be in an educational atmosphere unique to the United States in one or more of the following respects :

(1) Approximately 90% of all boys and girls enter secondary school, and more than 60% graduate.

(2) Except in the largest cities, secondary schools are ' comprehensive '—that is, they include university-bound students, those who intend to enter the family trade or business or who attend primarily for vocational training, and those who attend simply in consequence of school-leaving statutes.

(3) The sciences tend to be elective subjects. In the larger schools and the better small schools, courses in general science, biology, chemistry and physics may be available, but no student is required to take all of them, and in many schools he need take none.

Most commonly, one science course is required for graduation if the student is in the ' academic,' or university-bound group, since the universities require at least this much science for entry. It should be added that even in schools where a full battery of science courses is offered, it is most uncommon to make any linkage among them ; each science is presented as a self-contained one-year course. Except in the very largest cities, the secondary schools do not encourage specialized education.

The latest year for which figures are available in some detail is 1948–49, since the particular data in which we are interested are collected only about once in each decade. These showed that approximately one-third of the students were enrolled in either chemistry or physics during their secondary school years. Since chemistry ordinarily is favoured over physics by students in a ratio of about 3 : 2, it can be assumed that in 1948–49 about one in seven students studied physics ; there has since been no indication of any substantial change.

---

† Later papers, now being prepared for *Contemporary Physics*, will treat other aspects of the PSSC programme, such as films and teacher-training. Previously published in *Contemporary Physics* have been *A Survey of PSSC Publications* by J. L. Lewis (**1**, 248) and *The Summer Institute Courses* by H. F. Boulind (**1**, 309).

The typical physics student is a 16- or 17-year old boy in his last year of secondary school, whose academic standing places him in the top quarter of his class and who contemplates going on to higher education. Although boys and girls are almost equal, numerically, in the secondary school (49·2% boys, 50·8% girls) the ratio in the physics class is quite otherwise : boys usually outnumber girls by 8 to 1 or more.

All this descriptive material is relevant to an appreciation of the manner in which the PSSC course was developed. At the first full meeting of the committee, two important decisions were taken, reluctantly and yet (or so it seemed to the committee) necessarily. It was decided that the course developed by the PSSC should be intended for exactly such students as were then studying physics : a minority of the secondary-school population, comprising a portion of the academically-superior, university-bound students. It was also decided that the course developed by the PSSC should be a one-year course in physics, despite the belief that physics of itself warranted a two-year course, and that failing such a course in physics, a two-year course combining physics and chemistry would be far superior to two one-year courses, however excellent.

In making these hard decisions, the PSSC was moved by the realization that it was almost certainly beyond their abilities to insure adoption of any course that did extreme violence to current curricular practices. The American system of decentralized tax-supported education creates in effect a situation in which one must deal with a separate governing body for each small group of schools. Any attempt to make substantive changes in the general educational policy would merely prejudice whatever chances might exist of doing the essential job of improving physics instruction, and of putting it forth on its own terms. The PSSC limited itself, therefore, to the revision of the contemporary physics course, with the hope that such a course might of itself attract more students and a wider spectrum of student abilities.

These decisions narrowed the task of the committee to determining the content of a one-year course, and creating the texts and teaching aids in which that course was to be embodied.

So far as content was concerned, the heart of the charge brought against the conventional course was that it showed the ravages of time. So far as the secondary school was concerned, physics was primarily Newtonian mechanics to which had been added, more or less as patchwork, geometrical optics, heat, sound and electricity. Courses usually concluded with an obeisance to modern physics in the form of a few chapters of nuclear physics and electronics. It was a rare textbook which suggested that these ‘ units ’ were in fact part of the whole, and a rarer textbook still that made any real effort to show their interconnections.

Over the years, the texts had also become dominated by considerations of technology. It is not difficult to see how this had come about. Newtonian mechanics, which had once been mankind’s most enormous intellectual stimulus, has lost its immediacy in a world of fission and fusion and space adventure. If an ordinary student was to be aroused to an interest in the physics to which these texts introduced him, it was tempting to stress the cash value of the subject.

That, at least, may have been the theory, but in fact it had never quite worked out in practice, for the technology was usually out of date and superficial, and hence misleading.

Finally, the course had become overloaded. As new ' units ' were added, little attempt was made to eliminate or compress the old ones, and the technology had been heaped on top of all. No student could hope to handle, and no teacher to convey in any depth, the mass of material that was set forth. Instead, the teacher was constrained to require of his students that they memorize whatever salient facts were printed in bold-face type, and satisfied if they were able to repeat those facts with a minimum of distortion upon an examination paper.

As a consequence, physics has earned in the schools the reputation of being a dull course, thoroughly remote from the real world, and of value only insofar as it helped fulfil the admission requirements of the better universities. The student can hardly be blamed for this analysis, for judged on the basis of the physics he met in secondary school his conclusions were eminently correct.

This was the situation as it appeared to the PSSC when it became fully active in December 1956. By that time, preliminary discussions had taken place among a small group of physicists and educators, most of them at the Massachusetts Institute of Technology, Harvard University and Columbia University. The leading spirit in these early discussions was Professor Jerrold R. Zacharias of MIT, who, with Professor Francis L. Friedman of MIT, has continued to be the prime mover; he was firmly supported by Dr. James R. Killian and Dr. Julius A. Stratton, then president and vice-president respectively of MIT, and by Dr. Alan T. Waterman and Dr. Harry T. Kelly of the National Science Foundation.

The December meeting was attended by physicists and educators representing all sections of the United States. The meeting lasted for three days, during which were laid down the principles, and in general the programme, that have guided the work of the committee throughout its existence. The content of the course, as determined at the December meeting, is in its essentials the content that appears in the textbook. The materials which the December meeting called for have to a large extent been produced. The groups and individuals who volunteered, at that meeting, to play substantial rôles in the work of the committee have remained with the committee and contributed to its work.

The balance of this paper will be devoted to an account of the principles that have governed the choice and the treatment of the subject matter of the PSSC course, and of the manner in which these principles have been carried out.

2. Principles underlying the PSSC course

(1) The PSSC seeks to present physics as an intellectual activity, rather than as a body of rules for the control or the manipulation of natural phenomena. The course stresses the pursuit of knowledge and the human satisfaction that may be derived from the acquisition of knowledge, independent of the material gains that may be associated with it. It is quite clear that humanity, over the ages, has learned to use science to control its environment, but the physicist himself is moved rather by his urge to comprehend, and it is in this sort of spirit that the student is asked to approach the subject.

(2) The PSSC seeks to reflect in its course the spirit of inquiry. That the speed of light *in vacuo* may be stated to be so many kilometres per second is a significant fact and a useful one, but it is at least equally important that the student know why the speed has been measured, how it has been measured, what its significance and usefulness may be, and the reliability that he can reasonably impute to the measurement.

In the same spirit, the tentative nature of all physical knowledge is consistently stressed. Theory and experiment alike are presented as processes of successive approximation, in which final answers are neither expected nor sought. It is repeatedly made clear that any area of physics, however well-trodden, ends at last at a frontier beyond which lies the unknown.

In the laboratory programme, the student performs real experiments, in the sense that the answers are not known to him in advance. In so doing, he becomes acquainted at first-hand with the *dramatis personae* of physics, as it were. Once he has put waves through their paces on the surface of a ripple tank, and struggled with the phenomena that are revealed, the student is not likely to think of waves and wave motion as the remote abstractions of the physicist.

The laboratory is thus not intended to elicit for the student an inductive knowledge of the fundamental laws of physics ; even if this were possible, it would be uneconomical. Rather, his work in the laboratory should enable him to understand the raw material with which the laws deal and the manner in which the laws have been elucidated, and thus to participate at one or two removes in the achievements of the scientists who have erected, in the last few hundred years, the structure of science.

(3) The course reflects, as much as possible, the world of physics as it appears to the professional physicist, for only in this manner is it likely to have any clear relevance to the student himself. Because physics is a matter of great public interest in this era, the student is exposed constantly to information about what physicists are actually doing. Most of this information, as it passes through the various media of mass communication, is likely to be distorted out of all recognition, but it does establish the fact that physicists are not men who devote their lives to the calculation of the mechanical advantage of an inclined plane or the latent heat of vaporization of water. A student who must devote his classroom hours to such activities can hardly be blamed if he begins to feel that he has lost touch with the real world.

This does not mean that a secondary-school course in physics should deal exclusively with particle physics and quantum electrodynamics, and indeed any attempt to create such a course would clearly result in farce. It is possible, however, to direct a course in the general direction of modern physics, and to indicate at all stages that the student is indeed travelling a consequential road.

It follows also that the professional physicist must make himself responsible for the content of the course. This is by no means a small matter. It is possible to create a physics course which will reflect the preoccupation of the historian of science, or of the vocational-guidance counsellor, or even which will be designed primarily to fit the laboratory equipment which most schools now possess. Such courses in physics, and others, may in their own context be entirely worthwhile, but since it was never the intention of the PSSC to create any of them, the assumption of primary responsibility by the professional physicist is implicit in the PSSC programme.

The PSSC course prefers to present the traditional subdivisions of physics as various aspects of a single discipline. Thus the course lays great weight on the great principles of physics that cut across all lines of demarcation and unify the subject. As will be seen below, this resistance to compartmentalization has made it possible for the PSSC course to follow a general ' sortie-line ' intended to lead the student along a clearly defined structure from first chapter to last—a structure in which the subject of Light, for example, is not disposed of in one great gulp but reappears in progressively greater depth as the student's general understanding of physics deepens.

In more general form, the course also seeks to stress the interconnections among physics and the other sciences, and in particular those bearing upon chemistry and biology. Clearly, there is a limit to such a process in a one-year course, but the PSSC recognizes the desirability of the goal, and the course does what it can to satisfy it.

(5) The nature of the American school system makes it desirable to create a course which will be relatively independent of the order and content of the rest of the secondary-school curriculum. Many students will have studied chemistry before they elect a physics course, many will be offered chemistry after they complete a physics course, and the majority will not study chemistry at all. The PSSC course, therefore, must neither presuppose a chemistry course nor anticipate one.

The problem of the level of mathematics which the student can be expected to bring with him is more difficult, but again the nature of the school system makes it necessary to take very little for granted. Some students will study physics before they have encountered anything more than the most elementary algebra, others concurrently with a course in algebra or geometry, a few will enjoy a smattering of the calculus. The course therefore presupposes only a nodding acquaintance with algebra and a minimum of facility with its symbolism, and about as much acquaintance with geometry. Where more mathematics is required, as in some parts of Kinematics and Mechanics, it is introduced as part of the physics course; thus the solution of certain quadratic equations of motion are included in an end-of-the-chapter appendix.

(6) The most difficult decision which faced the PSSC at the outset of its work was the decision to omit from the course large areas of physics, and the selection of those areas which would be omitted. Similarly, the decision to minimize the amount of technology in the course was taken with some regret, in part because the connection between technology and physics is an intimate and significant one, and in part because to most students technology is an impressive by-product of science.

Consequently, PSSC has intended from the outset to provide for the student and for the teacher a means by which the course can be augmented in those areas where cuts have been made. As already reported in this journal, a series of books is being prepared to deal with subject matter omitted from the course or dealt with only in passing, as well as with the history of physics, the biographies of eminent physicists, and technology. These books, written by authorities at a level appropriate for an earnest student and available at modest prices, will hopefully go some way toward satisfying certain needs that the course arouses in the better student and does not itself satisfy.

3. AND THE ANTIPRINCIPLES

To the foregoing statements of what the committee thought it was obliged to accomplish, there must be added something of what the committee thought it should shun :

(1) The committee did not envisage its task as that of creating more scientists or more engineers. This, perhaps, is merely the obverse of the committee's belief that physics should be presented primarily as an intellectual activity, worthwhile in and of itself.

A course in physics then becomes as valid for the student who will go on to become an attorney, an entrepreneur or a politician as for the prospective scientist, doctor or engineer. For this reason, the committee never looked upon the course as being preparatory for further courses in a university, but rather dealt with it as if it were to be a terminal course in physics for most of the students who would elect it.

At the same time, the committee is made up of persons who have themselves responded to the attractions of a career in science or in engineering, or in the teaching of these subjects. As individuals, most of them undoubtedly would believe that a properly prepared and properly presented physics course will attract to similar careers many young men and women who are now repelled by the subject of physics, or by science in general, and who plan to shape their own lives in other directions. The course may very well, in the long run, increase the quantity and (more important) the quality of students who pursue the study of science and engineering. If this turns out to be so, the committee will be delighted ; at the same time the committee must be careful to consider this a by-product, and not the goal, of their efforts.

(2) The PSSC has sought neither to depreciate nor to diminish the rôle of the teacher in the learning process. On the contrary, the classroom teacher has played from the beginning an important part in the committee's programme. From the outset, secondary-school teachers have been important members of the PSSC staff, in numbers comparable to those of professional physicists, and one of the most carefully elaborated functions of the committee has been the accumulation of information and advice from teachers using the course. The material produced by the committee had always been intended to serve as tools by means of which the teacher might carry on his functions more efficiently, more economically and more fruitfully.

The shortage of physics teachers in the United States makes it inevitable that some schools will be badly served, and other schools not served at all. Ambitious attempts have been made to remedy this shortage by means of centralized teaching on television. The PSSC believes that this can be, at best, a measure of desperation, and that the ultimate solution lies in the better training of physics teachers, the provision of sufficient inducement to attract more capable men and women into physics teaching, and the provision of teaching aids which will increase the student–teacher ratio without affecting the quality of the teaching. The committee firmly believes that the application of technology as an aid to the classroom teacher has been badly neglected.

(3) It would be a denial of the entire philosophy of the PSSC to suggest that the course, as it now exists, is in any sense final. The text is complete only in that it is now being published between hard covers and will be generally available to the schools, but it is intended that it will be kept under careful scrutiny and

revised, when revision is warranted, in the light of pedagogical experience, or of relevant curriculum changes in other subjects, or of the growth of physics itself. A good deal of what has been wrong recently with the teaching of physics stems from the fact that it has become frozen into the mould of an earlier period. If in breaking that mould the committee were merely to create another, it would only encourage a recurrence of the situation it set out to correct.

It is true as well that the PSSC does not believe its solution to the problems of secondary-school physics to be the only possible solution or the best possible solution. Other courses, developed in the same spirit and with the same rigour as the PSSC course, will receive the enthusiastic support of the PSSC. It has always been the hope of PSSC that the completion of its own task would stimulate others to undertake similar undertakings ; certainly in a country as large as the United States, a diversity of course is to be encouraged and can be only salutary.

4. The form of the PSSC course

The first weeks introduce the basic concepts of time, space and mass not in terms of the thankless search for definition, but by demonstrating the manner in which they are handled. In the laboratory, the student measures time intervals of the order of 10^{-2} seconds by building his own stroboscope and measuring the period of vibration of an ordinary electric bell ; he proceeds to assemble a simple range-finder with which he measures distances in the range of several hundred metres, and an optical micrometer which takes him to the range of 10^{-5} metres.

While the student is carrying out these measurements, films extend the measurements of time and space intervals beyond the limits he can reach in the school laboratory. Great pains are taken to relate the films to text and laboratory. Thus the film entitled *Large Distances* builds from the basic idea of triangulation as exemplified for the student in the range-finder he has himself used ; he is shown how the same principle is used to map a continent and a solar system and finally, by using the orbit of the earth as base-line, to estimate distances to the nearest stars ; finally he is shown the line of reasoning that enables astronomers to estimate distances beyond the direct range of triangulation.

Although the apparatus used throughout this course is kept simple, understandable and inexpensive, this approach should not be taken as an intent to limit the student to crude instruments. On the contrary, he becomes acquainted both in film and text with precision instruments. In an early film, he is quickly introduced to such a device as the cathode ray oscilloscope, but he is led to it step-by-step, from the short time intervals that he himself has measured. The oscilloscope is presented frankly as a black box, but one which he can himself calibrate and thus intelligently trust.

As part of the study of motion he learns to plot distance, speed and acceleration curves for bodies moving in a straight line, and how to pass from one curve to another by graphical differentiation and integration. In the study of mass, with balances of his own construction, the student measures gravitational masses to the range of 10^{-4} grams.

The metric system is used throughout, and MKS units are standard throughout the text. The student is required also to familiarize himself quickly with power-of-ten notation, and with the notion of approximations to the nearest power of ten, for much of what he does later in the course relies upon this familiarity. It is driven home to him early that physics is not exclusively the search for another decimal place.

The student next encounters evidence for an atomic theory of matter, as exemplified in experiments he himself carries out with monomolecular layers and with a crude spectroscope, and in films dealing with crystals, the laws of definite and multiple proportions, and the behaviour of ideal gases. The last of these serves also as an introduction to the powerful concept of an idealized model.

The most recent revision of the text includes a section built upon experiments in which the radioactive disintegrations of various elements are counted. It is shown that such counting leads to the determination that the number of molecules is the same for equal volumes of two gases at standard pressures and temperatures. In this manner, the text arrives at Avogadro's number (being careful all the while to point out that the procedure being used is an historical inversion.) This provides entry on a brief excursion in chemistry, including a discussion of organic molecules and of Prout's hypothesis, and some brief descriptive material on isotopes.

It should be noted that an odd balance is sought in this portion of the course. The student enters the course equipped with a smattering of scientific knowledge, and certainly familiar with such matters as the existence of isotopes and the atomic nature of matter. It would be naïve to ignore this knowledge, and to pretend that he will learn of the existence of isotopes and atoms from the course. At the same time, the student ordinarily has no comprehension of the nature of the evidence for these beliefs, and others like them. To teach him demands a certain frankness on the part of the course and the teacher, as well as the introduction of such matters as isotopes long before the student would reach them in a course of which the order was dictated entirely by logical or historical considerations.

This terminates the first part of the course, and may be considered to be wholly introductory to the subject proper. For the most part, the student to this time has encountered familiar notions, although frequently in unfamiliar contexts and almost always in new depth. The material has been largely descriptive : the student has used a spectroscope without any hint of how and why it works. In all, he should have spent approximately one-fifth of his school year on these matters.

The student next enters upon the study of phenomenological Optics. The view of the majority of the committee, that Optics would provide a better point of entry for a serious study of physics than Newtonian mechanics, arose out of pedagogical considerations. Newtonian mechanics is highly abstract, and in the form in which it must be treated appears far too idealized to be a part of ordinary experience. At the elementary level, it is not particularly suitable to exemplify the interplay among experiment, law, and model that characterizes physics. In short, to the student who has not been prepared for it, Newtonian mechanics can appear hopelessly dull.

In Optics, on the other hand, the teacher can take advantage of phenomena with which the student is familiar, and about which his curiosity can readily be aroused. The rôle of the laboratory is clear-cut ; many experiments can be easily performed and made inherently interesting. The place of theory in physics is clearly exemplified. The teacher, with all of this in his favour, has some hope of building among his pupils a momentum that will carry them forward into mechanics, where he will find them better prepared to recognize the fundamental importance, the usefulness and perhaps even the beauty of this part of physics.

The Optics section begins, in the laboratory and the classroom, with observations of the behaviour of light, and with reflection and images, then proceeds to refraction and the behaviour of optical instruments. At this point, an attempt is made to put a model behind this behaviour ; the model that is first chosen is the Newtonian particle model, which is shown to account satisfactorily for most of the phenomena so far met. The model is used to predict that light will exert a pressure, and in a film the pressure of light is demonstrated. It is then shown that the Newtonian model predicts light should speed up when it changes direction toward the normal in passing from one medium to another. In a film, a measurement is made of the ratio of the speed of light in air to its speed in water, and it is shown that the observed experimental results are contrary to the prediction drawn from the model.

Now the notion of a wave is introduced, as the only imaginable alternative to a particle as a means of carrying ' something ' from place to place, and the student embarks upon the study of wave motion. This portion of the course is heavily dependent upon the laboratory, and in particular upon the ripple tank. To a great degree, it has been successful in the schools because PSSC was able to develop a ripple tank made of ordinary materials and reproducible for a few dollars (fig. 1). The tank is a far better teaching instrument than those which were previously available, in part because its low price makes it possible to supply a ripple tank for each group of three students or less. The motions of pulses are first studied, and then wavelength, frequency, refraction, dispersion, diffraction and interference are observed and measured.

This solid knowledge of the kinematics of waves leads naturally to the question of the behaviour of light again. The student has learned, by means of the ripple tank, more than enough to make him aware that light, if it is a wave, indicates by its gross behaviour that its wavelength length must be extremely short. He has learned further that if he is to find interference phenomena in light he must first find a means of resolving the problem of phase difference in differing light sources. These considerations lead directly to Young's experiment, where the student sees the interference, and is able to estimate the wavelength and to master a theory of interference in thin films and other phenomena of light.

This portion of the course occupies about eight weeks. Of the entire PSSC course, it is the section that relies most on the laboratory and least on films. In the schools, it has been spectacularly successful.

The transition from Optics and Waves to Mechanics is an explicit one. Having some knowledge of the manner in which things move, the student is ready to be convinced that he now requires some insight into the forces that lie behind the motions.

This portion of the course begins with Galileo's Law of Inertia and Newton's Second Law. (The course refers to this throughout as Newton's Law and has eliminated the conventional attempt to require the memorization of Newton's Three Laws of Motion.) Every attempt is made to allow the student to work out for himself, in the laboratory and before studying it in the text, the relationship between force, mass and acceleration. Rubber bands stretched a given amount are used as arbitrary units of force, bricks as arbitrary units of mass, 'frictionless' carts pull paper tape beneath the clapper of an electric-bell timer to measure acceleration (fig. 3). In the case of dry-ice pucks perfected by the committee, the inverted commas might almost be omitted ; Professor Edward M. Purcell of Harvard University has made two masterful films on Inertia and Inertial Mass with these pucks (fig. 4), and they are not beyond the means of a school to duplicate.

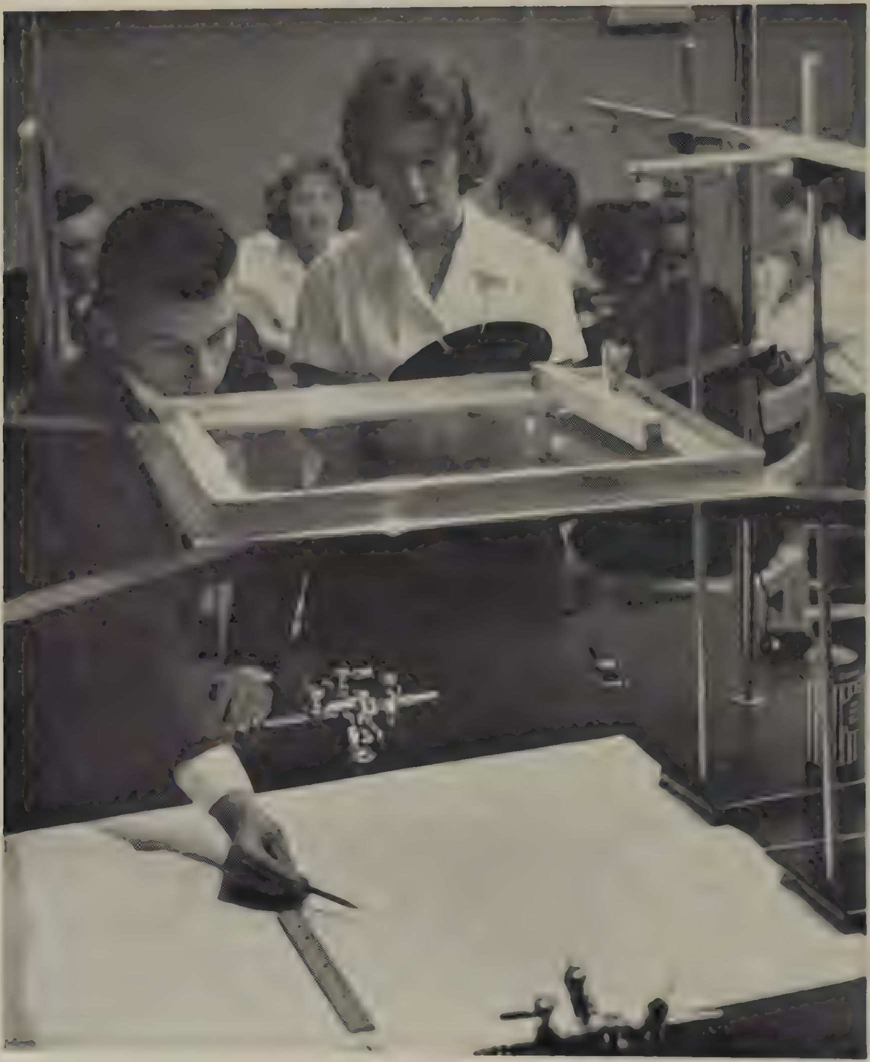


Fig. 1. Interference patterns being observed on the ripple tank.

Fig. 2. High school physics teachers study photographic techniques for experiments in dynamics at the University of Minnesota. Several of these teachers are teaching the new course in their schools this year as part of a nation-wide test of the programme.

Further experiments, performed in the laboratory, seen on film or described in the text, show the vector nature of the law, and describe projectile motion and the behaviour of a body under a force at right angles to its direction of motion, thus combining to establish the vector equation $\boldsymbol{F} = m\boldsymbol{a}$ as the basic law of Newtonian dynamics. The law is then used to develop the notion of simple harmonic motion. A film by Professor Donald M. Ivey and Professor J. P. H. Hume of the University of Toronto helps the laboratory work to distinguish between inertial and non-inertial frames of reference and demonstrates the fictitious forces that appear in a non-inertial frame ; this is considered also in connection with the diurnal rotation of the earth.

Upon the basis of all this, the Universal Law of Gravitation is shown to follow as a consequence of Brahe's observations and Kepler's analysis of them, and the student wins an acquaintance with both the solar system and with the means by which the solar system has been understood.

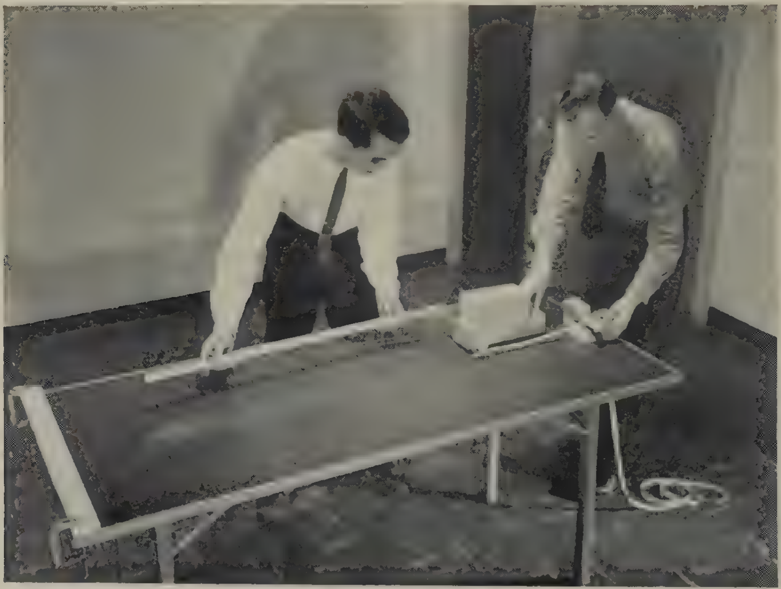
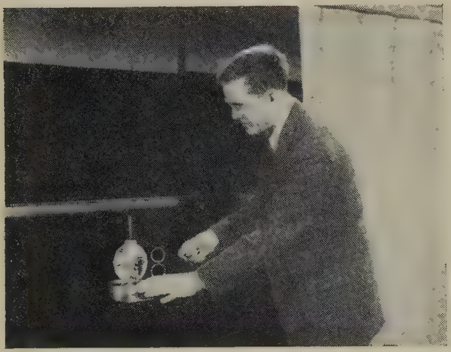


Fig. 3. Observing the acceleration of a ' frictionless ' cart. Force is being applied by an elastic band maintained at constant tension. The tuner is made from the capper of an ordinary electric bell; the paper tape is drawn under carbon paper upon which the clapper vibrates, and the spacing of the marks serves as a measure of the speed of the body which draws the tape.

Up to this point, the section has dealt exclusively with the force upon one body. Only now does he begin to deal with the forces on two bodies or on more than two. Momentum and the conservation of momentum are explored, again with the help of frictionless carts, dry-ice pucks and the electric-bell timer. The general principles are convincingly presented to the student, associated with the concept of the centre of mass and some considerations of the behaviour of rockets. The course then proceeds to the study of the notion of work, defined as a measure of energy transfer, and to the study of energy. Kinetic energy is first introduced, and shown to be conserved in elastic interactions. Potential energy is analysed in the laboratory by means of spring bumpers on carts, by the vertical vibration of a mass on the end of a spring, and the energy changes in a swinging pendulum. The student is able to discover that energy, considered as the sum of kinetic and potential energy, is conserved in these mechanical systems. Potential and kinetic energy in an inverse square force field (the earth's gravitational field) are next considered.

Next a simple gas system is introduced to associate temperature with molecular kinetic energy, and the student is led to the notions of internal energy, the equivalence of mechanical and thermal energy, and heat flow ; out of all this comes the complete principle of the conservation of energy, with which this portion of the course concludes.



Left: Fig. 4. A scene from the film *Inertia,* in which Professor Edward M. Purcell of Harvard University is demonstrating the behaviour of a dry-ice puck. Force is applied to the ice puck by rubber bands under standard tension; Professor Purcell is here preparing to show the effect of doubling the force.

Right: Fig. 5. Professor Francis L. Friedman of MIT demonstrating the apparatus used in the film *Millikan Experiment.*

Certain remarks come at once to mind in considering this section of the course. (1) The treatment is dynamical throughout, with statics treated, where it is treated at all, as a special case of the general dynamical theory. (2) Mechanics is treated inductively, with experiment and observation permitted to lead to the construction of general laws. (3) Heat is dealt with almost exclusively on the molecular level. (4) There is a stress on two-body interactions which is intended to prepare the student for considerations he will later meet in his exploration of the atom.

In this section of the course, as well as the section that follows, the intimate association of film and laboratory with the text is extremely valuable. The text, in addition to being the substratum of all the teaching, handles the burden of the mathematical treatment. The laboratory supplies the hard experimental evidence upon which the development of the subject matter relies. The films perform the basic experiments with a precision that is impossible in the school laboratory, and other experiments that are far beyond the resources of the best school laboratory, or indeed of any conventional laboratory. The association is in fact so intimate that special care must be taken in the management of films and laboratory, for a film shown out of place may weaken the effectiveness of the laboratory work.

The entire section is planned to occupy approximately ten weeks of the course. There is some evidence that this has been, up to now, a counsel of perfection, and that in practice it occupies substantially more. It may well be, however, that as teachers become more familiar with the course and have more experience with it they will be able to guide students through this section at a pace more nearly that anticipated by the committee.

From mechanical energy, the advance to a consideration of electrical energy is a natural one. Electricity is studied as the behaviour of elementary electric charges, static and in motion. The text and laboratory move quickly through the elementary considerations of electrostatics and the use of the electroscope as a point of departure for the study of electric currents; it is stressed that electrostatics and current electricity are merely two aspects of the same family of phenomena. An electric current is shown to be explicable in terms of the movement of electric particles. The battery is introduced as a chemical black box and extensively used as a source of constant emf. The conductivity of gases and solutions is demonstrated.

To account for conductivity in metals, the thermionic emission of charged particles is shown, and these particles identified with the negative particles postulated previously—the electrons. The text then briefly follows this trail to a consideration of diodes, electron guns and cathode ray oscilloscopes.

To this point the section has been entirely qualitative. Measurement begins in the laboratory and on film with the experimental evidence for Coulomb's Law, and its consequences are worked out. The Millikan experiment is performed on film, and in a subsequent film the experiment is scaled up by using plates of larger area and separation and an appropriate increase in the number of batteries; in this manner the student arrives at a value for Coulomb's constant and the force between two elementary charges. Finally it is shown that by means of an electron gun it can be established that an electron is associated with one elementary charge.

The concept of the potential as electric energy per elementary charge is introduced. The student studies the motion of electrons and hydrogen ions in a uniform force field, and learns the techniques by which the mass of each may be measured. The hydrogen ion is associated with the proton, and the planetary model of the neutral atom is presented.

On film and in text, the student studies electric fields. The volt is now defined in terms of one joule per coulomb, and the behaviour, although not the constitution, of an electric battery is once more described, this time in terms of the various units that have been set forth.

Electric current is now reintroduced as the transport of elementary charges in an electric field. Electric current is measured by electrolysis. The ampere is defined in terms of elementary charges per second, and the operation of a simple circuit is described in terms of the passage of elementary charges. The unit of power is introduced. Finally the theory established in the previous sections is extended to cover the relationship between potential differences and currents in gases, currents in electron streams boiled out of metals, and currents in the metals themselves.

The magnetic field, and the nature of its relationship to electric currents, are next taken up. Magnetic fields are measured in the laboratory, and their application to meters and motors touched upon. The forces on charged particles in a magnetic field, and the theory of the mass spectrograph, are discussed. The course deals next with induced currents and induced potential differences, Lenz's law and energy conservation, and finally self-induction and the electric oscillator.

Electromagnetic radiation is first approached by an appeal to symmetry : since a changing magnetic flux generates an induced electric field, it is asked whether the opposite is equally true. It is shown that if air in the gap between two charged parallel plates is ionized, a current passes between the plates but the expected magnetic field is absent ; from this it is reasoned that an equal and opposite field from another source has arisen to cancel the field due to the current. This second source is identified with the changing electric field between the plates. This phenomenon is now connected qualitatively with the generation of electromagnetic radiation. It is shown that accelerated electric charges radiate energy in the form of electromagnetic waves. And so the course returns to Light, and the earlier careful study of the atomic nature of matter, the kinematics of waves and the wave theory of light, and Newtonian dynamics begin to converge.

Armed now with a comprehension of waves, of Newtonian dynamics and of electrical forces, the student now possesses means to begin an exploration of the atom. He meets first, on film and by means of analogue models in the laboratory, the deflection of α-particles and the Rutherford model of the atom, and the experimental verification of Coulomb's law to intra-atomic dimensions. The Rutherford model quickly confronts him, however, with the inexplicable failure of an orbiting electron to radiate energy that this previous information would suggest.

This difficulty is compounded by further investigation into the nature of light. Experiments with the photoelectric effect lead to the Einstein photoelectric equation, which seems inconsistent with the wave picture that has so laboriously been built up. The photon interpretation of the photoelectric effect is presented, and it is shown that interference effects take place even within a system in which there exists but one photon at a time. The student is thus forced by the evidence to assert that light is neither wave nor particle, but somehow both: a particle that travels in accordance with the laws governing the behaviour of waves. Parallel experiments are introduced to show the wave characteristics of the electron, and the same conclusion is drawn. The concept of waves as patterns of probabilities is now inevitable.

In conclusion, the text and films look once more into the structure of matter. The Franck-Hertz experiment reveals the existence of energy levels within the atom, and associates them both with spectral lines and with photons. Further consideration of the spectrum lines leads to the notion of discrete energy levels associated with standing waves. The hypothesis of the standing wave is applied to the hydrogen atom, and shown to give satisfactory agreement. With this, the circle is closed : the waves of Part 2, the energy and the momentum of Part 3, the electrical forces of Part 4 are all bound into a single, coherent picture of matter and radiation—convertible into one another but with conservation of gravitational mass, inertial mass and energy.

It has perhaps been noted, or even anticipated, that in the latter stages of the course the reliance on film is progressively heavier, and on laboratory progressively less. This reflects the fact that schoolroom equipment for the sort of experiment required in modern physics does not exist. It is not a situation which the PSSC accepts complacently, and considerable work has been in progress during the lifetime of the committee to invent experiments in atomic physics which will be feasible in the secondary school. In particular, emphasis has been laid upon the development of a kit which will enable the student to build a variety of high vacuum tubes, and to carry out in this manner many of the experiments which now he may see only on film. Such a kit, designed by Owen Harries in Bermuda†, has recently been made available at a price which will bring the cost of a high vacuum tube to approximately $2. Its adoption will enable the student to do, rather than merely see, the experiments which make up the argument of much of the last part of the course.

To close this summary, it might be well to reiterate the PSSC's own view of what has been accomplished. The course, as it stands, seems to represent a reasonable satisfaction of the task that the committee set itself. It can certainly be improved in detail—but in general the committee believes that it meets the peculiar needs of the American school system, and that it constitutes a major advance over courses which are conventionally available in that school system.

It is too early to make any objective evaluation of the course in the light of experience in the classroom. What information the committee has gathered from approximately 1000 classes over a period of three years has been, on the whole, favourable. One test has been very satisfactorily met. Of 239 teachers who taught the course in 1958–1959 and who returned to teach physics during the following year, 228 or 96% elected to remain with the PSSC course rather than return to conventional courses. This is particularly heartening because it has been clear from the outset that the PSSC course, particularly during a period when the bulk of its teaching aids had not yet been available and laboratory material had not yet been made reasonably schoolproof, was an unconscionable strain on the teacher.

The committee believes that the PSSC course stands an excellent chance of surviving, that the clarification of ideas that has been a consequence of the labour of making the course is of permanent value, together with a great deal of the material produced. The success of the PSSC course has already effected great changes in the general outlook for the teaching of physics in the United States, and has stimulated work in other fields of science and in the humanities.

The Author:

Stephen White, who is in charge of its film section, joined the Physical Science Study Committee shortly after it was formed in 1956. He had previously written on physics for newspapers, the motion pictures, and television.

---

† *Editor's note:* This was described by J. H. Owen Harries, in a paper New High Vacuum Technique, in *Electronic Technology*, **37**, 8, 312, August 1960: it is also to be published in the *American Journal of Physics*, **28**, 8, November 1960.

Physics at the British Association, Cardiff Meeting, 1960

The high-light of the British Association Meeting at Cardiff was the Symposium on World Food and Population. This was remarkable because the several sections of the Association suspended their separate activities for the day, on September 5th, to assemble in the large Sophia Gardens Pavilion. The proceedings were extremely interesting to all scientists but, as they were not especially concerned with physics, they will only be briefly noted here. Physicists were, however, prominent in the symposium. Sir George Thomson took the chair at the morning session, and Professor P. M. S. Blackett summed up at the close of the afternoon session. It was clear that the members of the B.A. faced up to the fact that the present rate of growth of world population is dangerously large, and that even if tremendous advances take place in the use of land and even if the rate of production of food is enormously increased in the not too distant future, that growth is likely to lead to catastrophe. The position was ably presented in a nutshell by an eminent physicist who, in private conversation, commented that if a nuclear war broke out tomorrow and 200 million people were wiped out, then if the present rate of growth of population persisted, we should be back where we were within five years.

The Presidential Address was delivered by Sir George Thomson on ' The Two Aspects of Science ' in which he emphasised that science is not merely concerned with the control, but also with the understanding of nature, and he made a strong plea that the study of things of fundamental and permanent importance and the study of modern technologies should be held in equal honour. We would surely all agree with him that the return which pure science receives from industry in the shape of new instruments and new techniques is often not as gratefully recognized as it ought to be. The Address contained a masterly survey of many important facets of the quantum theory and its impact on the scientific thought of our time.

In recent years the Association has made great efforts to interest young people in its meetings, and it is fair to say that very definite signs that these efforts are bearing fruit were to be found at the Cardiff Meeting. The Association had provided Exhibitions to encourage our Universities to send young postgraduate students to the Annual Meetings, and these Exhibitions are often substantially increased in number by the Universities themselves. I took the chair at the meeting of Sub-Section A* on September 1st, when Professor J. L. B. Cooper of University College, Cardiff, gave an extremely interesting talk on ' The Main Lines of Mathematics ' and I was able to see that the majority of the members of his packed, enthusiastic audience could not have been over twenty-five years of age. It certainly was a decidedly more youthful audience than those which I found at Cardiff when I attended my first B.A. Meeting there in 1920.

The success of a Section Meeting depends to a very large extent on the ability and drive of the Recorder, and perhaps to a lesser extent on the President. On this occasion the President was Dr. R. van der Riet Woolley, The Astronomer Royal, and the Recorder was Miss M. V. Griffith, Head of Heating Research Division of the Electrical Research Association Laboratories, Leatherhead, Surrey, and both made notable contributions to the success of the session.

Dr. Woolley gave his Presidential Address on Stellar Motions. He pointed out that stellar motions relative to the sun are resolved into tangential and radial components. The former may be studied by comparison between new and old photographic plates of star positions, and converted into linear velocities if the star's distance is known. The radial component may be deduced from the Doppler shift in the spectral lines emitted by the star. Recent measurements by an independent method of the motion of cluster stars, such as the Hyades, have shown that their total motion relative to the sun is about 44 km/sec. He mentioned that, by discussion of the motion perpendicular to the galactic plane of nearby stars, a determination of the average density due to stars, dust and gas in the solar neighbourhood was possible. In a number of cases, stars wide apart in the sky appear to share a common motion, and when this can be established, the distances of these stars can be found, even if they are too far away for a reliable determination of trigronometric parallax; this was recently shown by Dr. Eggen at the Royal Greenwich Observatory.

Dr. Woolley dealt, too, with the study of the stellar motion of the Cepheid variables. Their average motion relative to the sun and its dispersion can be found from radial velocities alone. Their distances can be found from the period-luminosity relation, apart from a constant multiplier. If the latter is chosen to make the average proper motion agree with the average radial velocity, we may obtain the zero point of the period-luminosity relation uniting Cepheid variables, and so obtain the distance of the Large Magellanic Cloud and of other objects.

Dr. O. J. Eggen followed Dr. Woolley with an interesting paper on Stellar Groups, in which he pointed out that several groups of stars have been isolated on the basis of their motion, and showed that indicators of both age and abundance of heavy elements confirmed the hypothesis that the stars of these groups are coeval. Dr. R. D. Davies then discussed measurements of the magnetic field of the Galaxy carried out by a radio method at Jodrell Bank. The results were compared with relevant optical data and their importance for an understanding of the galactic structure was emphasized.

It would have been surprising if mention of the International Geophysical Year had not been made in Section A. Dr. R. Stoneley, President-elect of the Section for the Norwich Meeting in 1961, introduced the three distinguished contributors on September 1st. The first, Professor W. J. G. Beynon of University College, Aberystwyth, described some of the preliminary results of the I.G.Y. ionospheric programme, with special reference to British contributions thereto. The second, Mr. J. MacDowell, described some aspects of the geomagnetic and seismic observations made at Halley Bay, in the course of which he described a novel observatory built on a thick, floating ice-shelf formed of accumulated and compacted snow. He gave a fascinating account of the way in which seismographs were mounted under such strange conditions.

Finally, Dr. G. E. R. Deacon of the National Institute of Oceanography, dealt with I.G.Y. Studies of seismic and storm surges, changes in mean sea level and the general circulation of the water in the oceans. For example, he mentioned that near the British Isles the mean sea level is about 15 to 20 cm higher in autumn than it is in spring; while most of this difference in height can be accounted for by changes in temperature and pressure, there is some evidence that there is less water in the ocean at the end of winter. The mean sea level around southern England has risen by nearly 7·5 cm in the last 50 years and this rate of rise is now important in connection with the practical problems of coastal protection.

Dr. Deacon mentioned, too, the effects of a group of long waves sent out from a disturbance on the sea floor. Travelling at about 400 to 550 miles per hour in deep water, not more than 1 or 2 ft high, and with as much as 100 miles between successive crests, they cause catastrophe when they move into shallow water where their amplitude is greatly increased. He also described Dr. Swallow's ingenious float made from aluminium scaffold tubing. Sealed at the end, and loaded with batteries and a simple circuit with magnetostriction oscillator to produce necessary sound signals, the float gains buoyancy as it descends in the ocean, and may be loaded to float at any predetermined depth. In this way the movements of currents at great depths have been shown to be faster and more variable than was previously supposed.

The September 6th session was devoted to applied physics. Mr. W. I. J. Price of the Road Research Laboratory described electrical heating systems designed to prevent the sudden formation of layers of ice and snow in dangerous zones on roads, paying attention to the different types of ice and their relation to the heating control systems.

Dr. M. R. Hopkins gave an extraordinarily good survey of modern theoretical aspects of metal physics and discussed their industrial applications in the electrical and steel industries, while Mr. J. D. Murray of the Applied Metallurgy Section of United Steel Companies Ltd., gave an account of creep in steel. In Section A* an interesting discussion on the selection and training of computer programming staff took place.

Applied physics was also in evidence at the last session on Wednesday, September 7th, when Mr. I. J. Ley of the Food and Medical Section, Wantage Radiation Laboratory, discussed problems of food preservation and the sterilization of medical products by ionizing radiations, paying particular attention to effects on nutritive value and the possible causation of toxic effects.

Miss M. V. Griffith showed some very interesting experiments in thermoelectricity to illustrate her remarks on thermoelectric refrigeration. Based on the existence of the large Peltier effect at a junction between two semiconducting alloys, the method is of great theoretical interest, and has practical importance in the cooling of small volumes of material.

The final Wednesday at former British Association Meetings has always seemed to me to be of the nature of an anticlimax. But at Cardiff the many members who attended the last meeting of Section A were well rewarded. Indeed, some might have thought that without it, the physics content of the programme as a whole was somewhat meagre. Of course, the loss of a whole day occasioned by the general Symposium was exceptional. Incidentally, the old feeling of anticlimax was also dispelled in another way. The induction of the new President of the Association, Sir Wilfrid Le Gros Clark, contrary to previous custom, took place later in the morning and provided an opportuniy for the thanks of the Association to be expressed to the Civic Authorities of Cardiff in a dignified manner.

As is usual at British Association Meetings there were many interesting excursions. Among them were visits to Berkeley Nuclear Power Station; The Steel Company of Wales at Port Talbot; Standard Telephones and Cables

Ltd. at Newport, Mon., and British Nylon Spinners at Pontypool; there were also visits to the Physics Departments of Cardiff and Swansea. A good programme of scientific films was arranged and many members took the opportunity of visiting the Science in Schools Exhibition held in the New Buildings of the Welsh College of Advanced Technology. This was arranged by schools in the Cardiff, Newport and Monmouthshire areas, to illustrate the work in Science and allied subjects by pupils in secondary schools. I did not manage to visit this Exhibition, which was open at times inconvenient for me, but I did hear favourable comment about the physics experiments, particularly those from secondary modern schools. Each afternoon in the Great Hall of the Welsh College of Advanced Technology a lecture for young people was delivered, and among them was one on Atomic Radiation and Materials, by Professor A. Charlesby.

Forty years ago one could almost certainly expect to hear for the first time an announcement in Section A of some great new advance in experimental physics from a British or overseas physicist. Because of modern facilities for immediate publication of results of interest, and because of considerations of priority, this state of affairs no longer obtains. But it is fortunately still possible for a young worker to hear and see great men and to enjoy their wider surveys of physics at the B.A.

L. F. BATES.

The Nuclear Physics Meeting of the Institute of Physics and the Physical Society, Glasgow, September, 1960.

by W. S. C. WILLIAMS
Department of Natural Philosophy, University of Glasgow

On 14th and 15th September the Department of Natural Philosophy at the University of Glasgow was the host establishment to the second 1960 nuclear physics conference. These conferences are held twice a year, previously under the auspices of the Physical Society, now under the auspices of the combined Institute and Society. The previous meeting had been held at Liverpool University and a report by Dr. Green and Dr. Muirhead has appeared in the June 1960 issue of *Contemporary Physics*. These authors have described the aim of these meetings, and the intention that younger research workers should participate actively has been upheld. Of thirty-three short contributions presented at Glasgow, about one-third were given by postgraduate students.

The conference opened on the morning of the 14th with a speech of welcome from Professor Dee. Professor Dee then handed over to Sir John Cockcroft who, as President, took the chair at the business meeting of the Institute and Society; and then in turn invited Professor Rochester to take the chair at the first session.

The first session entitled ' High Energy Physics ' was opened by Professor Burhop who reviewed work recently presented at the Rochester Conference on the strange particles. He mentioned work at Berkeley consistent with charge independence, and an experiment designed to determine the relative parity of the Λ and Σ hyperons. This involves searching for discontinuities in the cross-section versus energy variation for the reaction $\pi^- + p \to K^\circ + \Lambda^\circ$ at the threshold for the reaction $\pi^- + p \to K^+ + \Sigma^-$. Discontinuities in the angular distribution coefficients were found; these would suggest even relative parity, but the statistics were too poor to make a definite conclusion. A review paper by Professor Kemmer followed. He did not confine his subject to the high energy field, but traced the history and development of the theory of elementary particles and their interactions.

The success of the CERN proton synchrotron has been very much in the news and Dr. Wetherell reviewed the latest results obtained during investigations on the secondary particle yields from internal targets. We already knew that positive particle yields had included deuterons at momenta up to 6 GeV/c at 16° from targets bombarded with 25 GeV protons. This had been a very surprising result for although the incident energy is very high, one does not expect a loosely bound object such as the deuteron to be ejected from a struck nucleus with energies of 5 GeV and such large transverse momentum. Wetherell reported that the beam was found to contain tritons and He^3 nuclei also. All these particles have sizes of the order of 10^{-12} cm whilst their observed energy corresponds to momentum transfers which can only occur at impact distances of the order of 10^{-14} cm. However, it is possible to think of reactions in which an incident proton creates a nucleon-antinucleon pair in the field of a target nucleon. The final state contains at least three nucleons which must be very close together just after pair creation and it is not impossible that they recoil bound. For example $p + n \to d + p + \bar{p}$ or $p + n \to H_e{}^3 + \bar{p}$.

It is interesting to note that no high energy α-particles have been observed (Cocconi *et al.*, 1960). To produce these in a simple two nucleon interaction would require the production of at least two nucleon-antinucleon pairs, which is very much more unlikely. Thus the observations are in favour of the recoil from anti-nucleon production rather than effects associated with the disruption of nuclear matter, which could be expected to yield α-particles. In parallel with exploration, CERN has been setting up beams for experimental use: examples of these are the 25 GeV proton beam and a 16 GeV π^- beam. These have been used for total cross-section measurements and the results of this work were reported in two short contributions in Session I. A hydrogen bubble chamber has been exposed to the π^- beam and the 50 000 photographs are being analysed by groups at CERN and five European Universities. This is a very large undertaking but the results are beginning to appear and were presented by Dr. Merrison in a short paper. He drew attention to the observation that hyperons produced by 16 GeV π^- went predominantly in the backward direction (referred to centre-of-mass coordinates). This is not surprising as it corresponds to the direction of lowest momentum transfer to the baryon. However, one would expect the K° mesons to reflect this distribution in the forward direction, whereas, in fact, they are produced almost isotropically. However there is probably a considerable $K^\circ - \bar{K}^\circ$ pair production which could overwhelm any anisotropy.

The meeting was due to divide after tea into Sessions II and III, Experimental Techniques and Theoretical Topics. A review paper by Dr. Whitehead from AERE on Solid State Detectors had a natural place in Session II but it was considered to be of interest to all the delegates and it was therefore read at the end of Session I. These devices depend for their action upon the collection of electron-hole pairs produced by the passage of a charged particle in the depletion layer which exists at, for example, a *p–n* junction in germanium or silicon. By appropriate techniques the junction can be generated at the surface of a semiconductor and it is possible to measure the energy of particle stopping in the layer with very high resolution and linearity. At this stage of development the maximum thickness of depletion layer which can be made corresponds to that which would stop 12 MeV protons. These devices will be very valuable in low energy nuclear physics and it is expected that development will continue and reveal new areas of application.

The writer attended Session II on Experimental Techniques: it consisted of short contributions. Dr. Manning, from AERE, described a measurement of the *g*-factor for free protons at 150 MeV which confirms the theoretical predictions concerning the magnetic precession of polarization vectors; also, it gives an upper limit on the electric dipole moment of the proton. The current interest in the μ-meson was reflected in the description by Dr. Michaelis of a beam transport system matched to the CERN synchrocyclotron and which is designed to produce intense and pure beams of μ-mesons. These will be used in scattering and absorption experiments. Other short contributions concerned bubble densities, neon flash tube techniques and a π^+ detection technique.

The second day of the Conference opened with Session IV, π- and μ-meson physics. Professor Hamilton reviewed the π-meson physics. He reported some interesting work on direct and indirect methods of measuring the π° lifetime. The value is still very uncertain but about 10^{-16} seconds is favoured.

He also discussed the interesting resonances in the $\pi-p$ interaction at about 600 and 900 MeV. These seem to occur in the $D_{3/2}$ and $F_{5/2}$ states of the system but there remain some difficulties in reconciling this with the observed scattering angular distributions at these energies. Professor Hamilton also dealt with the hypothetical $\pi-\pi$ resonance. This is now a very fashionable concept with the theorists but the experimental evidence is very meagre. The difficulty lies in separating the $\pi-\pi$ resonance effects from those of the π-nucleon interaction. He also reported an interesting suggestion due to Drell which concerns meson photoproduction at very high energies. A meson pair is produced in a nucleon field; one of the pair is virtual and transfers four-momentum to the recoiling nucleons. At very high energies this four-momentum approaches that characteristic of a real meson and the cross-section for production of the real meson fo the pair is considerably enhanced. If these ideas are correct very high energy electron accelerators will prove to be more copious suppliers of mesons than proton accelerators of the same energy. This argument is giving the promoters of the Stanford 45 GeV electron linear accelerator considerable confidence.

Dr. Muirhead followed with a review of μ-meson physics. He stressed the similarity between the μ-meson and the electron; they have the same charge, spin, and g-factor and they appear to have the same interaction with other particles. He reported work on the nuclear capture of negative μ-mesons. Recent evidence that there was some anomalous large angle scattering of μ-mesons in lead, which has been the subject of much research, he found to be unproven. In our experience the μ-meson seems to behave like a heavy electron.

There were six short contributions in Session IV. Of particular interest was one of the reports from Durham University describing the measurement of the μ-meson spectrum at sea level and the calculation, from this observation, of the spectrum of π-mesons which generates these μ-mesons. The interesting problem which arises is that this derived spectrum differs significantly from that observed directly by Bristol University in high altitude emulsion work.

The Conference proper ended with Session V on nucleon forces and related problems. Dr. Willis reviewed the methods of calculating the binding energies of the light nuclei and Dr. Phillips discussed the two-nucleon problem from the phenomenological viewpoint. He stressed that the phase-shift analysis was far from complete and that much more experimental data on nucleon-nucleon scattering was required.

No account of the Conference will be complete without mention of the other activities. Most important of these was the Rutherford Lecture which was held on the evening of the first day: this is an annual event promoted by the Physical Society, to be given by eminent scientists working in those fields which have their origins in the work and discoveries of Lord Rutherford. Glasgow had the honour of hearing Professor C. F. Powell give the eleventh lecture in this series and his title was " Nuclear Interactions at Very High Energies ". Sir John Cockcroft introduced Professor Powell with some reminiscences of the Professor and his interaction with Cambridge. He told of his return to Cambridge to attend the *viva-voce* examination for his Ph.D.: it was necessary to stay a night in his old college but he found sleep impossible for the night was bitter, the heating non-existent, and the bedcover inadequate. Not wishing to be exhausted before his examination the next day Professor Powell was at last forced to lift the linoleum from the floor and to lay it upon the bed in order to provide himself with extra thermal insulation. Professor Powell started his lecture by adding to this story. He had once heard C. T. R. Wilson extolling the virtues of newspaper (particularly *The Scotsman*) for keeping the body warm when observing clouds from mountain tops. He admitted that on the occasion to which Sir John Cockcroft referred he had added to his bedclothes all the newspaper lining from the drawers in his room before resorting to the linoleum.

Professor Powell's lecture dealt mainly with the results of recent work at Bristol University exposing photographic emulsion to cosmic radiation at very high altitudes. He discussed the π-meson spectrum and the discrepancy between the Durham derived spectrum and that observed by his own group. He also described another phenomenon which is as yet unexplained: most of the very high energy ($>10^{12}$ ev) showers, generated via neutral π° meson production, develop according to theory but occasionally showers are observed which, after normal development and degeneration, start to redevelop as if supplied with more energy. Many explanations have been examined but none fits the observed development. These showers that Professor Powell's group are studying are sometimes so energetic that the number of shower particles becomes exceedingly large and the emulsion so darkened that not only is it impossible to identify individual particle tracks but the photographic effect of the shower becomes visible to the unaided eye. Professor Powell showed slides of such photographic records and concluded his lecture with a ciné film showing the launching of balloons carrying the emulsion to the required height for exposure.

There were several social activities arranged in connection with the Conference. Delegates were invited to a reception given by the University on the evening of the first day and this was followed by a dinner in honour of Professor Powell. His lecture followed this dinner. The host establishment, the Department of Natural Philosophy, was ' at home ' to delegates in the evening after the Conference. The Glasgow University 330 Mev electron synchrotron and HT sets were on show. The evening was completed by refreshment of the most acceptable kind. The day after the Conference was devoted to a steamer trip to the Kyles of Bute and Tighnabruaich. As far as is known all delegates returned safely to Glasgow.

The next Conference in this series will be at the University of Birmingham during the Easter vacation of 1961.

REFERENCES

V. T. Cocconi, T. Fazzini, G. Fidecaro, M. Legros, N. H. Lipman and A. W. Merrison, *Phys. Rev. Lett.*, **5**, 19, (1960).

The Author:

Dr. W. S. C. Williams is lecturer in the Department of Natural Philosophy in the University of Glasgow. His work is in the field of high-energy nuclear physics, and he is at present working on the nuclear scattering of high-energy x-rays.

Opening of the Atomic Energy Establishment, Winfrith

On Friday 16th September 1960, The Atomic Energy Establishment at Winfrith, Dorset, was officially opened by the Rt. Hon. Viscount Chandos, P.C., M.C; it had, in fact, been in operation for more than a year.

1. THE WORK AT WINFRITH

The Establishment is part of the Research Group of the Atomic Energy Authority and its work consists of—

(a) *Reactor physics*—investigations into the way in which the nuclear chain reaction is maintained by the flow of neutrons in a reactor. The work is partly theoretical and partly practical. The experimental work mostly consists in assembling fissile material, with or without a moderator, and making a chain reaction take place at a very low rate so that the heat output is negligible. Many important features of the behaviour of a reactor operating at a high power level can be deduced from these studies.

(b) *Basic reactor engineering technology*—work concerned with the flow of heat in nuclear reactors, with the flow of fluids which are to remove the heat and with associated phenomena. Although there is nothing novel about the use of fluids to transport heat, the reactor designer requires information that has not so far been found necessary for conventional engineering purposes.

(c) *Control and instrumentation*—the theoretical study of the general problem of controlling reactors.

(d) *Computing services*—the use of the digital and the analogue-type electronic computing machines.

(e) *Medical services and radiological safeguards*—radiation monitoring services to safeguard against exposure to radiation including environmental surveys of the area surrounding the site.

There are a number of types of nuclear reactor under development, namely—

Magnox reactors. Reactors of this type, using magnesium alloy sheathing, have been operating since 1956 at Calder Hall and since 1959 at Chapelcross. Larger reactors of this type, of more advanced design are being constructed, and at Winfrith the reactor physics effects of the build up of plutonium in the uranium fuel are being studied. [SCORPIO I and II.]

Advanced gas-cooled reactor. This reactor is based on the use of uranium oxide sheathed in stainless steel or beryllium. The moderator is to be graphite and the coolant carbon dioxide. The rate at which heat is to be extracted from the fuel is to be higher than that which can be attained by the magnox type. The fuel will be slightly enriched uranium-235. At Winfrith research in the reactor physics and also on the stability of the thin sheaths of stainless steel is going on.

High temperature reactor. In this reactor there is envisaged the use of refractory non-metallic fuel consisting of fissile material dispersed in a solid moderator. The fuel will contain thorium which will be transmuted to a fissile isotope uranium-233, so maintaining reactivity. Helium is to be used for a coolant because of its chemical inertness.

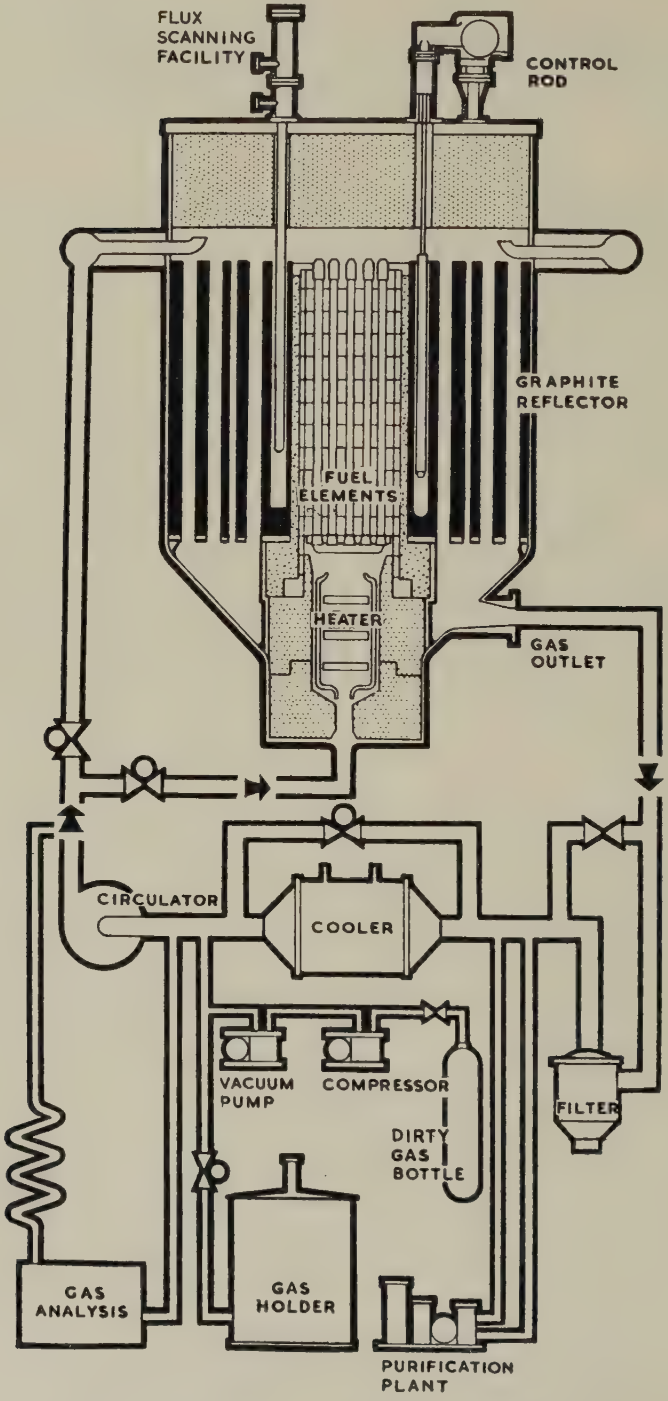


Fig. 1. Construction of ZENITH, showing the reactor vessel, and the nitrogen circulation and purification system.

Fast reactors. In these reactors no moderator is used so that the neutrons retain the high velocities with which they emerge from fission. The coolant is liquid sodium or sodium/potassium alloy. A zero energy fast reactor—ZEBRA—is to be built at Winfrith for studies connected with the development of this system.

Liquid moderated reactor systems. In these reactors the functions of coolant and moderator can be combined and the use of liquids containing light or heavy hydrogen is envisaged [HELEN]. Among these types are the boiling water reactor and the organic liquid moderated reactor. Work is going on in connection with the basic engineering technology of heat transfer in the systems including studies of boiling.

2. Zenith

The ZENITH reactor, which was the first to be built at Winfrith, started operation for the first time in December 1959.

ZENITH operates under temperature conditions similar to those expected in a high temperature gas-cooled power reactor. The core consists of a nearly cylindrical assembly of 235 fuel elements 8 ft high and 4 ft in diameter, in the form of graphite cylinders, containing pellets of fissile uranium oxide, fertile thorium oxide and graphite, arranged in regular patterns. The ends of the fuel elements are made of solid graphite which, with a 3 ft thick graphite annulus surrounding the fuel, form a reflector to return some of the neutrons escaping from the fuel into the core. The core is enclosed in a pressure vessel through which nitrogen is pumped in a continuous closed circuit. The nitrogen is heated by a 200 kw heater and the hot gas raises the core mean temperature to 800°c. At the top of the core the nitrogen is mixed with cooler gas and then passes through the reflector which it maintains at a mean temperature of 400°c. It is important to maintain the purity of the nitrogen and so a purification system is used to remove traces of carbon monoxide and carbon dioxide. The reactor pressure vessel and associated plant are enclosed in an air-tight building. ZENITH is operated at a peak neutron flux of 10^8 neutrons/cm^2/sec at a maximum nuclear power of 100 w; the nuclear heat produced is therefore negligible in comparison with the energy used to heat the core.

This reactor, of which the cost including design, construction and commissioning was approximately £400 000 is being used by the international staff of the OEEC High Temperature Reactor Project to establish design characteristics for a 20 MW high temperature gas-cooled reactor experiment called DRAGON which is being built at Winfrith.

3. Dragon

The OEEC High Temperature Reactor Project (DRAGON) is a joint undertaking by the U.K., Austria, Denmark, Norway, Sweden, Switzerland and the European Atomic Energy Community, Euratom (Belgium, France, Italy, Luxembourg, the Netherlands and West Germany). The agreement provides for a joint research and development project over a five-year period from 1st April 1959.

The objectives of the DRAGON project are to carry out a programme of research and development on the high temperature gas-cooled system, and to

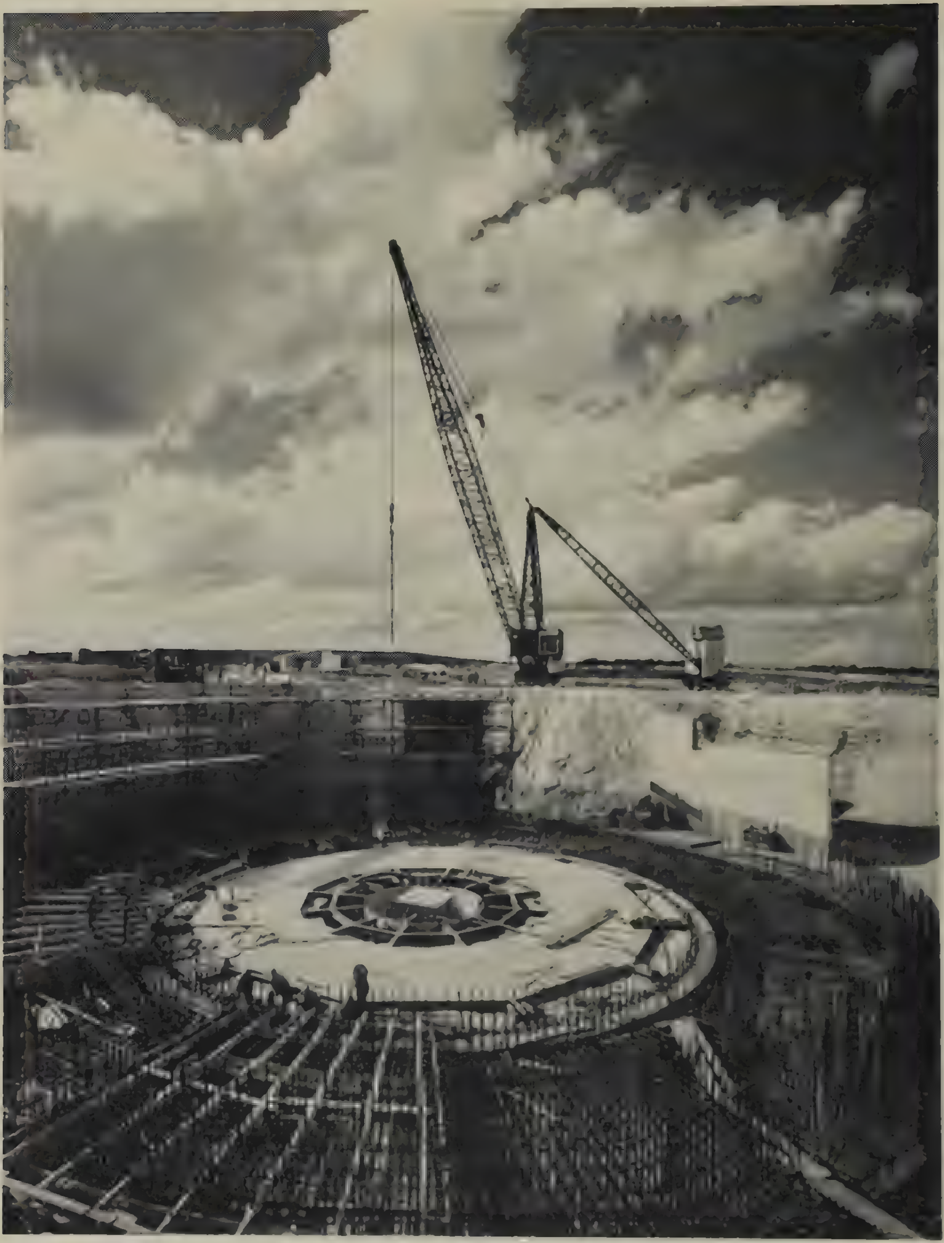


Fig. 2. Reinforced concrete foundations for the DRAGON Reactor Project.

design, construct and operate an experimental reactor in order to obtain practical proof of the feasibility of some of the novel features which this type will incorporate. The system chosen for study employs graphite as the moderator and helium at 20 atm pressure at the coolant. The fuel is a mixture of uranium, thorium and graphite. The use of part of the graphite moderator as the fuel container, in place of the more usual metal cans, is a major advance. Its success will depend on obtaining a sufficiently impermeable graphite to avoid excessive activity in the coolant. The programme of work on both graphite and fuel element inserts will involve irradiation testing in research reactors in this country and in Europe. In parts of the core temperatures may reach 1500°c.

4. Hector

Early in September 1960, the news was released of the signing of the construction contract for another reactor at Winfrith. The reactor, which will be called HECTOR (*H*eated *E*xperimental *C*arbon-in-the-form-of-graphite *T*hermal *O*scillator *R*eactor), has been designed to facilitate the use of the ‘ Pile Oscillator ’ technique (which makes use of small samples), for evaluating the nuclear behaviour of fuels, moderators, and structural materials in reactors. The reactor core is an annulus of graphite and highly enriched uranium/aluminium alloy within which various assemblies of fuel and moderator may be constructed around a central tube. The sample under investigation will be cycled through this tube, producing oscillations in reactor power level from which the nuclear behaviour of the sample, under the conditions reproduced by the assembly, can be determined. Since the fuel material in the assembly may be natural or enriched uranium-235, plutonium, thorium, uranium-233 or irradiated elements from power stations and since the assembly can be heated to 450°c it will be possible to investigate samples under the conditions to be found throughout the life of many different designs of power reactors. Successful measurements are possible only if the temperature of the heated assembly is extremely stable; it is desirable that the moderator temperature should not change by more than 3×10^{-3} °c/min.

5. The ancillary services

The visitors invited to Winfrith, as well as seeing the monsters, and enjoying a conducted tour into the bowels of ZENITH, were allowed to see some of the ancillary research serving the monsters. Amongst a considerable number of exhibits the writer found the following of exceptional interest.

Studies of organic coolants. The exhibit showed the apparatus used to measure the physical and heat transfer properties of the polyphenyls, which are relatively resistant to thermal and radiolytic breakdown; these organic liquids are attractive for use as coolants as they are non-corrosive and have low vapour pressures.

Boiling water. The various phenomena associated with the formation of steam bubbles on heated surfaces and the flow of these bubbles in water are extremely complex. The exhibit illustrated the various modes of boiling that occur and the determination of the associated heat transfer performances, flow characteristics and void distributions of two-phase mixtures.

Lifting gear for charge-and-discharge. An open mock-up of the DRAGON reactor experiment was shown in which the charge-and-discharge machine was working. Consider the precision of adjustment demanded for the grabbing, lifting and withdrawal of these long rod loads! Nearby there was a model—about 4 ft high—showing the details and all the operations and motions of the final scheme; this was a quite fascinating piece of ‘ meccano ’ type assembly.

Bearing tests. On account of the unusual operating conditions which will be found in DRAGON no protective oxide films will be formed on the surfaces of the bearings in the reactor experiment and so investigations into bearings' performances are in hand. The exhibits showed a number of rigs for carrying out plain and ball-bearing tests in a pure helium atmosphere at high temperature.

Pure helium. Diagrams of the equipment for the purification of helium were shown and explained.

Fluid flow measuring instrument. This instrument is based on the effect produced on a magnetic field by a flowing polarized non-conductor. This method is of special interest as no solid impedes the motion of the liquid and also because it can be used with liquids which may be already radioactive. (The demonstrator informed the writer that this piece of apparatus will, in due course, be shown at a Physical Society Exhibition.)

Differentially heated graphite assembly. Local temperature variations occur within the lattice of an operating power reactor. These variations affect the physics of the reactor and lead to changes in power level. The assembly shown had been constructed to provide basic data concerning this effect. The assembly consists of a large block of graphite within which tubes of various materials, notably graphite and beryllium, are heated electrically. Neutrons provided from antimony-beryllium sources mounted below the block diffuse through the assembly. The neutron characteristics within the tubes are measured and from these results the effect of heated regions of moderating material can be assessed.

Other exhibits which the writer heard of but missed seeing were an ionization chamber making use of a.c.; and some new methods of measurement of temperature and the new materials used in them—this investigation is necessary because standard thermocouples cannot be used in the DRAGON reactor core since they would react with hot graphite or would not withstand the intense level of irradiation.

Also housed at Winfrith are a joint team from the UKAEA and the industrial Consortia which is studying the physics of graphite moderated gas cooled reactor systems under the *B*ritish *I*ndustries *C*ollaborative *E*xponential Programme—the BICEP Agreement, and a study team from the British Shipbuilding Research Association which is examining the possibilities and problems of the nuclear propulsion of merchant ships.

V. T. SAUNDERS.

International Conference on Physics Education

UNESCO House, Paris, July 28–August 4, 1960

This Conference, under the auspices of OEEC and the International Union of Pure and Applied Physics, was attended by more than a hundred delegates, representing some twenty-six nations. The Conference Chairman was Professor Sanborn C. Brown (U.S.A.), and the Conference Secretary Mr. Norman Clarke. The forty papers to be discussed had been circulated to delegates well before the Conference, and the actual sessions were devoted to general discussions based on groups of papers which formed a unifying theme. Immediately after the close of the Conference, Prof. Brown and Mr. Clarke began the preparation of a full report of the proceedings, which will make a substantial volume. Among the general topics discussed were the place of physics in general education (opened by Mr. Clarke); the selection of university entrants for physics courses (Prof. G. K. T. Conn); examinations in physics (Dr. H. F. Boulind); the use of films and television in physics teaching (Dr. Harvey White, U.S.A.); and the Physical Science Study Committee's syllabus and programme (Prof. J. R. Zacharias, U.S.A.).

Paper No. 40, *The Teaching of Physics in Schools*, the report of the working group set up by the Office for Scientific and Technical Personnel of the Organization for European Economic Cooperation (OEEC), under the chairmanship of Mr. Clarke, was the first to be discussed in the opening session on physics in general education. This paper urges the need for a presentation of physics in terms of modern ideas of the subject rather than, shall we say, in the way that the people who taught our more elderly teachers learned it. It is easier, of course, to draft a syllabus than to make it work in the classroom and laboratory, and some difficulties were raised in the discussion; but the problem will have to be faced. But let Paper 40 speak for itself. In the opening paragraph there is the following: " Physics is not a collection of ' facts ' which can be learned; it is a highly imaginative intellectual structure of concepts that gives a meaningful and creative picture or model of such of man's experience of the world in which he lives as it has yet been possible to integrate into a consistent whole. It is a picture that is constantly being given further detail and some parts of which are occasionally being withdrawn". The paper moves on to consider the purpose of school education and arrives at the conclusion—" The purpose of school education is to prepare children for the adult world . . . . and . . . . few subjects can compare in importance with science as a necessary foundation for understanding that world ".

It is worth while to consider this statement side by side with extracts from other papers, viz. from Paper 15, *The Teaching of Physics in the Age Group 15 to 18 Years* by MM. Dechene and Eurin: " Teaching in this age group is the teaching of thought structure rather than the teaching of facts. In this teaching, physics, which allies knowledge of the concrete to mathematical reasoning and which permits the coordination of a very large number of facts by a very small number of ideas, must have a high place " and from Paper 9, *Physics in General Education* by Yoshinobu Kakiuchi (University of Tokyo): " If we are to meet the real purposes of general education the student should be trained to behave rationally, responding to rapidly changing outer circumstances physical as well as social. In order to behave rationally knowledge about our external world is first of all required . . . . human knowledge can never attain its final completion. Even a single future experience might require the revision of our present conception of the world as was the case of the wave theory of light at the time of discovery of the photo-electric effect. This means that no single item of knowledge about the external world can be of absolute validity independent of its empirical background ". In Paper 9 there is also the challenging sentence " Traditional education often lays too much emphasis on the logical aspect of things ".

Paper 40 mentions defects of present physics courses—" . . . . the most serious criticism of traditional methods of teaching physics is that it is only in the latest stages in the course, often only in the stages that would not be taken by the children who leave school at sixteen, that there would appear to be any relationship between physics as taught and physics as even the most obtuse member of a modern community must be vaguely aware of it ".

A small but carefully chosen exhibition of teaching apparatus and demonstration experiments was held in the foyer of the UNESCO building during the whole of the Conference. The commercial exhibits by the teaching-apparatus firms were enterprising, and in many cases rather costly by any standards; but there were also some excellent kits of elementary apparatus. Instrument firms who have been giving their minds to the educational uses of their products made some useful contributions. The simple but effective apparatus brought over by Prof. Zacharias showed that fundamental experiments can be done well without a lot of fuss and expense. And the teaching and conversazione experiments from British and French university departments were a delight.

About fifty films were exhibited, including distinguished examples from the Advanced Science series of the Educational Foundation for Visual Aids, and a number of excellent filmed lessons produced by the Physical Science Study Committee, the American Association of Physics Teachers, and Encyclopaedia Britannica Films.

At the closing session of the Conference, a number of resolutions were accepted without dissent. These, briefly summarized, were:—

(1) That physics should be regarded as a necessary part of the education of all children.

(2) That an international committee of professional physicists should be established to devise means for improving the teaching of physics.

(3) That the Conference Chairman should ask the General Assembly of the International Union of Pure and Applied Physics to accept this responsibility.

(4) That efficient teaching of physics needs specialist teachers who can keep abreast of developments in a rapidly growing subject; and that Governments and universities should consider some general steps (set out in detail) for improving both the efficiency and the attractiveness of physics teaching.

(5) That further international conferences should be held from time to time to examine and discuss the problems of physics education.

Conference of the Physical Sciences Section of The Association of Teachers in Colleges and Departments of Education

The Physical Sciences Section of the Association of Teachers in Colleges and Departments of Education met at Furzedown College, London on October 1st to consider the problems facing the Teacher Training Colleges in their work of preparing students as specialist teachers of science, competent to teach to at least the O level of the G.C.E. Examination.

The preliminary discussion on policy turned largely on the recruitment of students to the advanced courses which, with new laboratories and expanded staffs, the Colleges were providing. It was felt that Heads, Careers Masters and Mistresses and the Science teachers of the Grammar Schools *had little or no knowledge about these courses.* It was agreed, therefore, that an approach should be made to the appropriate bodies so that a better knowledge of the opportunities presented by the Training College science courses could be disseminated amongst the schools.

Difficulties in the recruitment of suitable staff demanded publicity within the University Honours Schools as well as the Training Departments. In addition the provision of simple research facilities, together with close liaison with the appropriate University department, was essential if the quality of the science teaching in the Colleges was to be maintained and further improved.

Specialist discussion on syllabuses in physics and chemistry revealed that science students in the Colleges were receiving an average of 10 time-table periods weekly on the sciences ; that the entrants to the courses lacked experience of practical work and, whilst competent in their ability to answer A level examination questions, lacked a full understanding of the fundamentals of their subjects. It was agreed that syllabuses should be orientated towards modern work and towards a concentric view of the subject. It was further agreed to set up a working party to consider syllabuses in physics and chemistry which would be suitable to the needs of the present day.

During the afternoon, Professor G. F. J. Garlick, Professor of Physics in the University of Hull, spoke on " The Education of Physicists in the Universities ", and considered the particular problems of the provincial Universities. The candidates differed markedly from the ideal for they were already the finished products of the schools, where the effects of the teacher shortage were often all too apparent. There was not a great deal to choose between those with good A-level passes, but possessors of poor A-level marks had to be treated with reserve. On his arrival in the Honours School, it was essential to put the fresher on his toes and to help him to make the difficult transition from being taught to learning how to study for himself. To this end a personal supervisor encouraged him to put his thoughts on paper. This slowed down the all too prevalent rapid scanning technique of reading and assisted in the process of divorcing the student from his inherent belief in the infallibility of the text-book.

In the first year it was essential to reinforce the foundations and to present physics as a whole. Usually great weaknesses were found in fundamental matters, especially in dynamics and electricity. Linear motion, rotation and vibration should all be presented at the same time, for a good knowledge of vibrations was a primary need of the physicist. The mechanics needed was not " applied mathematics " but an integrated view of dynamics.

In electricity the atomic concept should be introduced as soon as possible—and it was to be hoped that schools might eventually rid themselves of the concept of the isolated pole which caused students to have such difficulties with that of the dipole. It was necessary to give students the ability to think in terms of atomic particles, whilst some appreciation of the concept of field potential, whether gravitational or electric, was very desirable.

Together with this first year study of mechanics, optics, (especially wave motion of all types), thermodynamics and kinetic theory were studied. The first year course concluded with an historical account of the bases of the atomic theory which stressed the transition of thought from the classical to the modern.

During the second year an intensive course of atomic physics (mainly concerned with the outer atom) followed in which the behaviour of the particles was represented in modern terms, rather than through the historical development.

The most difficult year to organize was the third. Recognizing that no undergraduate could be brought to the frontiers of the subject in three years, it was necessary to decide what to select. Difficulties of comprehension made it essential to provide a more than adequate course in Wave Mechanics and, because the unity of physics is so clearly shown in these fields, in the physics of the nucleus and the solid state. It was also necessary to keep the number of lecture hours low and to space them suitably. During this year fewer experiments were done, but selected students were given special problems to solve in association with the research department. This provided a great stimulus and was of far greater value diagnostically than the traditional six hour practical examination which could well be abolished.

The assessment of the student was an affair involving the human element—and his supervisors of each of the three years of the course were involved in it. A thorough knowledge of the student as a person was therefore available to his assessors.

The graduate of such a course knows where to go next and how to learn. But to assist that further learning, some formal post-graduate teaching (say two lectures a week) was necessary.

In conclusion, Professor Garlick reminded us that the tone of the undergraduate school is set by minds on the borders of the subject—by the members of the research school in fact. The maintenance of a strong research tradition is essential to the healthy Honours school. The " bulge " must not be allowed to smother this tradition.

E. J. WENHAM

Contemporary Reading

The two publications that were mentioned at the end of the article in the June issue must be dealt with first. *A Physics Anthology*, edited by Norman Clarke, published for the Institute of Physics by Chapman and Hall (London) and Reinhold (New York) at 35*s.* is a miscellany of contributions to the Institute's *Bulletin* that I greatly enjoyed re-reading. They include Alan Bullock on *Science and the Humanities*; Sir Cyril Hinshelwood on *Physics among the Sciences*; Herbert Dingle on *The Origin of Modern Science*; K. A. G. Mendelssohn on *Max Planck*; A. C. Crombie on *Newton's Conception of Scientific Method*; F. I. G. Rawlins on *Physics and Aesthetics*; R. V. Jones on *The Theory of Practical Joking*; N. R. Hanson on *The History and Philosophy of Science in the Undergraduate Physics Course*; and a symposium on *Ethics and the Scientist* presided over by the High Master of Manchester Grammar School. There is also a number of more or less technical papers on educational practice. It should be appreciated by physicists, and welcomed also by the general reader who wants to understand the physicist's point of view; it may give the layman the answer to the question " Are physicists human? ", and dispel any lurking doubts we may have been harbouring ourselves.

I said also that I would review fully the Annual Report of the National Physical Laboratory; but fortunately a number of contributions from members of the staff have since been promised, and it would seem preferable now to leave things to them.

The Sixth Annual Report of the United Kingdom Atomic Energy Authority (H.M.S.O., 5*s.*) gives the usual Select Committee Report, Balance Sheet, and account of the production plants. Chapter IV, on the Reactor Development Programme, describes the studies made on the properties of graphite, the development of beryllium and stainless steel as canning materials for the advanced gas-cooled reactor (AGR), the start-up and teething troubles of the Dounreay fast breeder reactor, and the high-temperature gas-cooled reactor (HTGC) project which may be a major source of power in the 1970's, with progress on the OEEC Dragon programme at Winfrith to test the principles and materials. In general research and development, experiments on nuclear-nucleon scattering with beams of polarized protons, on the properties of 14 MeV neutrons, and on the interactions of neutrons with fast-reactor materials, have been made using a number of new devices which include the neutron booster of the Neutron Project Laboratory and the 12 MeV tandem Van de Graaf. Basic research on the gravitational ' red-shift ' using the narrow-width γ-radiation from iron-57, and work on controlled thermonuclear reactions and the properties of plasma both at Harwell and in industry (the Intermediate Current Stability Experiment, ICSE) are described in Chapter VI. Chapter VII mentions briefly methods foreshadowed for the monitoring of nuclear weapon tests, including any that might be exploded in outer space. Chapter IX deals with health and safety measures, and notes the special postgraduate courses in biophysics at several universities that have already followed the interim recommendations of the Veale Committee. An illustrated summary of the report, called *Atom* 1960 (H.M.S.O., 2*s.* 6*d.*) is published for the general reader; this is very good value.

The fifth edition of Roberts and Miller's *Heat and Thermodynamics* (Blackie & Son, Ltd., 1960, 45*s.*) incorporates some additions to the text, and a number of references, covering the work of the last decade. Adiabatic demagnetization taken down to nuclear-spin level (pp. 158-169), the triple point of water (p. 341), the gas turbine and gas refrigerator (Chapter XIV) and superconductivity (Chapter XVII) seem to be the chief additions, but the revision has permeated the book. There is a good deal, for example, on zero-point energy which was not in the earlier editions.

Modern Physics, by M. S. Smith, the senior physics master at King's College School, Wimbledon (Longman's, 12*s.* 6*d.*) is intended for undergraduates and the good sixth-form students. After two chapters on the concepts of physics and the classical picture, he deals with electrons and X-rays, ions and atoms, motion, relativity, the quantum theory, matter waves, wave mechanics, and the nucleus and nuclear theory. It is an excellent book, and deserves to step into the almost unique position that the early editions of Crowther's

'Ions, Electrons, and Ionizing Radiations' occupied about forty years ago. What distinguishes it from the routine exposition of atomic physics is that the author knows exactly where he is going *and assumes that the reader also is a person of industry and purpose.* He says " a good knowledge of physics up to A-level, a maturity of outlook, and a willingness to think about the logical and philosophical basis of the subject are taken, for granted ".

Equally demanding are the authors, G. S. Christiansen and Paul H. Garrett (Connecticut College) of *Structure and Change: An Introduction to the Science of Matter* (W. H. Freeman & Co., San Francisco and London, 63*s*.). They ask of the reader " an open mind, some tough thinking, and a fair share of hard work ". This offers an integrated course embodying physics, chemistry, astronomy, and cosmology, starting from " where the student now stands —that is, from usual commonsense information and ideas of a partially educated non-scientist ". The book is planned in three parts, (1) *The Mechanical View of the Physical World*, (2) *The Electrical View of the Physical World*, (3) *The Modern View of the Physical World*. Some of the chapters are quite brilliant—notably that on *The Mechanical View of the Earth* —and the third part of the book is up-to-date and most instructive. It is well illustrated with original drawings in two colours; there is a useful bibliography at the end of each chapter, as well as a selection of new and searching exercises. It is attractive and intelligently conceived, but I would suggest that the reader who will get most out of it is the person who already knows a good deal about the subject-matter and is primarily interested in pursuing the authors' synthesis of physical science as a united whole. I find it difficult to believe that a beginner could master it, but if it is really intended for somebody like an arts graduate who has some acquaintance with scientific ideas and with the historical background, the contents are appropriate. But wait. Half-way through the book one finds: " It should certainly be evident by this time, but it often needs emphasis, that an atom and an ion of the same element are quite different substances. The two entities have different formulas. It is of paramount importance to distinguish between them ", and in the appendices, instruction on the use of algebraic symbols and the meaning of direct proportion. It is illuminating to read that the official view of the Government of South Africa in 1890 was that the earth is flat and covered by a heavenly dome, and that Captain Joshua Slocum had to be explained away to President Paul Kruger; but marginally relevant in the context. My own diagnosis would be that, in trying to bring an appreciation of modern science down to levels where both literacy and numeracy are unawakened, the authors have produced a text which is repetitive and rather uneven to read. In short, while it deserves every success, and will certainly be approved by reviewers who go through it at one swoop and appreciate the hard work and ideas behind it, it is difficult to visualize it, say, in class use. It might, of course, be a useful book in the hands (or on the bookshelf) of a teacher with a gift for précis.

G. R. NOAKES.

ESSAY REVIEW

La Structure et l'Évolution de l'Univers. Reports and Discussions at the Eleventh Solvay Conference. (In English.) Edited by R. Stoops. (Brussels: 1958.)

The eleventh Solvay Conference on Physics, held in Brussels from 9 to 13 July, 1958, has given rise to this interesting and stimulating volume. Part I of the book, dealing with general cosmology, begins with a paper by G. Lemaitre on " The Primaeval Atom Hypothesis and the problem of the Clusters of Galaxies ". Lemaitre first points out the importance of the question: can the whole spatial universe be covered by a finite number of finite regions of space? Only if we believe that it can will it seem worthwhile to attempt to draw inferences about the structure of the whole universe from observations of a finite part of it, since otherwise the cumulative errors of extrapolation from observations of only limited accuracy, made in the finite part of space accessible to us, will deprive statements about the universe as a whole of observational meaning. Similar considerations apply to the space-time universe.

The first space-time universes sought for were static ones, in which the large-scale properties of space remained constant. Physicists did not wish to depart farther than was necessary from well-tried concepts, only recently found inadequate. After the discovery of the red-shift, solutions were sought in which the only change with time was a change of scale. The universe has a beginning, but is always at relatively the same distance from it.

A different type of approach seems natural, says Lemaitre, if one aims at finally constructing a theory which shall include quantum phenomena. In the quantum theory, we treat any physical system as an assembly of ' states ' which may or may not be ' occupied '. Mathematical equations, characterizing the system, make it possible to evaluate the relative probabilities of the different states as deductions from an observation made on the system. They also predict the probability of transition from any one state to any other. We take as model of an expanding universe a physical system in which the number of states increases with time. The number of states at the beginning of the model universe is to be as few as is conceivable. As time goes on, the total energy of the model universe is split into smaller and smaller packets, which may fill the different states. The mode of splitting, and the course pursued by the evolving universe, are only probabilistically specified. Space and time begin to take shape as the number of states becomes very large.

According to this view, widely different universes could evolve from the same beginning, but once the number of states has become sufficiently large the further course of large-scale events is nearly deterministic, in consequence of the statistical properties of large aggregates. The beginning of this model universe is called by Lemaitre the Primaeval Atom. He goes on to consider cosmological processes, assuming that at a very early stage the universe consisted of an assembly of corpuscular rays shooting outwards in an expanding space which originally was filled by the Primaeval Atom. The highly speculative introduction serves merely to give an idea of the world-picture motivating this assumption. The topics briefly considered in relation to this model universe include cosmic rays, formation of galaxies, the cosmical term in the relativistic equations, and clusters of galaxies. A short paper by O. Klein follows, which deals in a more mathematical way with questions relevant to the early development of the system of galaxies.

The next report is by F. Hoyle on the steady state theory. T. Gold collaborated in the final sections. Hoyle introduces the formalism of the theory as a generalization of the ordinary hydrodynamic equations of a fluid and proceeds to develop its consequences, comparing them from time to time with observational data and with the predictions of relativistic cosmogony. The treatment is both simple and profound, but it is hardly possible to summarize it here. A short quotation will give some idea of its scope: " It may be added that some restriction can be placed on the extent to which the natural constants of microscopic physics can have changed during the past. Thus a rather complicated quantity built out of the gravitational constant and the constants of microscopic physics appears to approximately the seventh power in the expression for the luminosities of main-sequence stars. In view of this sensitive dependence it would not be possible to change the quantity in question by any appreciable factor without running into serious conflict with the astrophysics of stellar evolution ".

" Other restrictions arise in connection with the study of geophysics and of meteorites, particularly concerning the half lives of naturally radioactive elements. If the natural constants have varied appreciably over the last five billion years, it is necessary for instance that they should have varied in such a way as to prevent serious discrepancies between widely different methods of estimating the ages of meteorites, e.g. the strontium-rubidium method depending on β-decay and the uranium-lead method depending on α-decay."

A stimulating paper from the philosophical point of view is T. Gold's contribution " The Arrow of Time ". Gold starts from a consideration of the old question: how is it possible that we are able to tell the direction of the flow of time in a universe where all the physical laws are symmetrical with respect to a reversal of the time direction? That we can do so is usually considered to be implicit in the well-known theorem that " the entropy of an isolated system will always increase and never decrease " as time elapses.

It seems strange that a statistical discussion of the system—namely the consideration of its entropy—should be able to tell us something—namely the direction of ' time's arrow '—which could not be inferred from a detailed calculation of the behaviour of the system according to the laws of physics. Why does the arrow appear when we are dealing with the statistical superposition of effects, each of which is determined by laws which have no arrow? Gold concludes that it is the expansion of the universe which, in the last analysis, provides the necessary asymmetry. For it is this which enables the depths of space to act as a ' sink ' for radiation. But this means that we are bound to see the universe as expanding; the time direction in which the universe is expanding (not contracting) is the one in which radiation ' gets lost ' in space, and this is the direction from past to future.

In the discussion, Prof. L. Rosenfeld makes the point that the question of time-direction arises at a deeper level than the statistical; namely in connection with the object-observer relation. The transmission of a signal from object to observer already defines a sense of time, and the universe might be observed as either expanding or contracting by an observer whose time-direction is fixed in this way. The question whether it is meaningful to separate an observer from the universe does not seem to have been raised.

Part I ends with a report by seven authors on possible observational tests of Einstein's general relativity theory that relate to the structure and evolution of the universe, and with a paper on World Models by E. Schücking and O. Heckmann.

Part II of the book, entitled " Survey of Experimental Data on the Universe ", consists of two reports with accompanying short discussions. The first, by J. H. Oort, is a survey of present knowledge about the distribution of galaxies and of mass density in the universe. The second, by A. C. B. Lovell, gives an account of radio-astronomical observations which may give information on its structure. Professor Lovell's general conclusion is that, as yet, there are no radio observations which can influence significantly the existing views on the large-scale structure of the universe. Of course, many highly interesting results of a more localized character have been obtained (some of these are described in Part III), but until the problem of the unidentified radio sources is cleared up it seems unsafe to draw any general cosmological conclusions from the radio observations.

The third part of the book deals with the evolution of stars and galaxies from the observational side. It begins with a report by H. C. van de Hulst on radio investigations of the galactic system (i.e. our own spiral nebula, one of whose stars is the sun) and nearby galaxies, and on the evidence for magnetic fields in galaxies. One of the most interesting of the recent results mentioned is the probable existence of expanding spiral arms near the centre of our own galaxy. Observations with the 25-metre radio telescope at Dwingeloo indicate a continuous mass of gas (briefly, an arm) between us and the centre, moving away from the centre with a velocity of about 50 km/sec. If this is regarded as a one-way outward motion, it implies an enormous outpouring of energy from the central nucleus of the galaxy. There is no known source for this energy at present. If we supose that the motion is really a two-way one, it becomes necessary to explain how the gas can remain invisible to radio observation during its return journey. Van de Hulst thinks that a one-way outward motion maintained by enormous non-gravitational forces seems at present the more plausible explanation of the observations.

The evidence for the existence of large-scale magnetic fields in the galaxy is rather indirect, but very few astrophysicists doubt their existence. It was first suggested as a means of keeping the high-energy cosmic-ray particles from escaping from the galactic system, consisting of galaxy and galactic halo. (If they could escape, it would be necessary to suppose the whole of intergalactic space filled with cosmic ray particles, and this seemed a larger assumption than the other.) Polarization effects observed in starlight later provided supporting evidence for the existence of the magnetic fields; several thousand stars have been measured and the polarization values are similar over fairly large areas. The polarization could result from the passage of the starlight through interstellar dust (known to be present in abundance in the galaxy) in which the more elongated dust particles have been partially lined up by the effects of an interstellar magnetic field. Chandrasekhar and Fermi have suggested that the spiral arms are tubes of magnetic force, but there is no observational evidence sufficient to give a decision of this question at present. Non-thermal radio emission is also observed; and is generally supposed to arise from a synchrotron mechanism similar to that in the Crab nebula, namely the motion of ' relativistic ' (high velocity) electrons in a magnetic field. The measurement of the Zeeman effect in the 21 cm line is mentioned at the end of the report as a hopeful possibility.

V. A. Ambartsumian's report on the evolution of galaxies is especially concerned with the study of clusters of galaxies and the inferences about galactic evolution which can be drawn from this study. In multiple galaxies, the period of revolution may be of the order of a billion years (10^9) and the age of the galaxies themselves is probably of the same order. Thus multiple galaxies, and clusters of galaxies, may be expected still to show some traces of the primary conditions under which they were formed. This is confirmed by the high percentage of ' trapezium type ' configurations in multiple galaxies; that is, configurations containing at least three members at roughly the same order of distance from each other. Such configurations would be rare in multiple galaxies which had had time to make more than a few revolutions since their formation. Again, in some large clusters of galaxies the relative velocities are so great that they must represent disintegrating systems. The proximity of their members indicates that the clusters are being observed at an early stage in their life-history.

If we assume that some clusters and groups of galaxies consist of recently formed components rapidly flying apart, we may ask whether the components originate from a single amorphous mass, or whether the primary nucleus of a galaxy has divided into separate parts which give birth to independent galaxies of the cluster.

The division of the nucleus and separation of its parts in the middle of an already existing galaxy would cause very violent processes lasting for some tens of millions of years. The new nuclei can be imagined (says Ambartsumian) to eject matter which by spreading forms rapidly growing shells consisting of gas and stars—young galaxies moving through the primary galaxy. In the radio galaxies Cygnus A and Perseus A we may possibly have such a situation. In the latter we see what looks like two galaxies moving with a relative velocity of at least 3000 km/sec. In the former we observe two nuclei within the same galaxy. Both are very intense radio sources. The Perseus A source has commonly been interpreted as two colliding galaxies. But if collisions were the explanation of these radio-emitting pairs, we should expect (says Prof. Ambartsumian) to find among them many more dwarf galaxies than we do. Actually, all known galaxies emitting exceptionally intense radio-waves are super-giants. The provisional conclusion is that Perseus A and Cygnus A are systems in which the nuclei have been divided but the separation of galaxies is not yet complete. The blue ejections from the central parts of certain elliptical nebulae may be very young galaxies. The bridges and wispy filaments linking many double and triple galaxies are, on the above theory, probably the last bonds still linking galaxies that have separated by division of the nucleus. In some cases (for example NGC 7752–7753) the link is an actual spiral arm, rather than a thin filament.

In the discussion which followed, Dr. Oort expressed doubt as to whether the evidence was sufficiently compelling to justify the adoption of such a revolutionary idea as the fission of galaxies. Professor McCrea quoted A. Salam as saying that nuclear physicists would not be surprised if nuclear interactions of a certain order of ' weakness ' (technical sense) could create new matter.

There are two more contributions, one by Hoyle on the origin of the elements in stars, and one by W. W. Morgan on spectroscopic investigation of the relative numbers of ' old ' and ' young ' stars in galaxies. A brief general discussion concludes the book.

E. H. Linfoot.

BOOK REVIEWS

The Special Theory of Relativity. By J. Aharoni. (Clarendon Press: Oxford University Press, 1959.) [Pp. viii + 285.] 45*s*.

This is a scholarly work which offers useful reading matter for serious students of mathematical physics. It is, however, only fair to say at the outset that this book is too erudite and too highly specialized for the main stream of Physics Honours students, for whom there are some simpler and cheaper monographs.

The approach adopted in this book is worthy of comment. Relativity and quantum theory were, in their early stages of development, largely independent. To achieve a fusion of the two, new techniques have to be developed and new concepts emerge. The author has undertaken the task of building the mathematical apparatus needed for this modern exposition and little or no previous knowledge of tensors, matrices and group theory is required of the reader.

Starting with the concepts of space and time in inertial frames, with the usual clocks and rods hurtling through space, the author introduces tensors and then applies the apparatus to Maxwell's electromagnetic theory and to general field theory. Relativistic particle dynamics is followed as a sequel by relativistic hydrodynamics, but whereas the dynamics chapter is of great interest to the physicist, especially for its discussions of conservation laws and of the electron in an electromagnetic field, the implications to the physicist of the hydrodynamics chapter are not immediately apparent.

The final chapter on spinors, some 87 pages in length, presents an introduction to a rather difficult subject, developed to enable the quantum mechanical entity *spin* to be incorporated in relativistic field theory. This is very heavy going and the ultimate objective never seems to be clearly defined.

Except for an excessive use of the first person plural, the style is pleasing and the generalized treatment of the various topics will appeal to mathematicians. It is, however, claimed that the book is intended for physicists. There is indeed plenty to stimulate their appetites but the fare is rather unsubstantial. This arises because the author is apparently unprepared to compromise with the crude assumptions and approximations which are so necessary if one is to apply to real physical situations a theory which is not yet fully developed. There are several examples of this kind where the physicists may be frustrated by the vagueness of the conclusions and the absence of an intelligible summing-up after working through an argument which at first seemed destined to get to grips with some noteworthy problem. It occurs with the clock paradox, where an orthodox treatment by use of special relativity has as sequel a passage on the application of general relativity, marred by ambiguous phraseology which may leave the reader quite baffled. There are similar but even more regrettable anticlimaxes in the discussions on beta-decay, on mesons and on parity.

Again, most physicists would like to have more information on the experimental foundations on which the theory rests. This should include some numerical values and a clear critical account of the observational method. It is an important part of the training of a physicist to develop his critical faculty and the author might well play his part in this process by spending more effort on the foundations of the edifice than on the mathematical building operations whereby the extensive and beautiful theory is elaborated. Most of what the author has to offer in this way is relegated to the Appendix.

There are very few literature references in the main text, but a list of references is given at the end of the book. The book is excellently produced, in the manner and high standard one expects of its publisher.

A. J. Woodall,
C. G. Wilson.

Non-Linear Wave Mechanics. By L. de Broglie. (Elsevier, 1960.) [Pp. 304.] 57*s*.

In his *Scientific Autobiography*, Planck records that the Energetics controversy gave him " . . . an opportunity to learn a fact—a remarkable one, in my opinion. A new scientific truth does not triumph by convincing its opponents and making them see the light, but because its opponents eventually die, and a new generation grows up that is familiar with it ". One is reminded of this fact when one realises how far de Broglie's point of view (that beneath the nowadays familiar framework of quantum mechanics with the probability interpretation of the wave function there should lie a fully deterministic physical world) is removed from the general trend of thought of modern physics. Apart from a few of de Broglie's own pupils, there are practically no younger physicists who doubt the intrinsic truth of the philosophy developed by the Copenhagen school under Bohr. We may all be wrong, and it is therefore extremely illuminating to read de Broglie's account of various attempts made by himself to obtain a causal, non-probability interpretation of quantum mechanics. The book is well written and should give food for thought, even if it is doubtful whether it will convince any opponents. D. TER HAAR.

Physics of the Upper Atmosphere. Edited by J. A. Ratcliffe. (New York: Academic Press, 1960.) [Pp. xi + 586.] $14·50.

The study of the atmosphere at heights above about 60 km demands knowledge in many branches of physics, and there is much to be said in favour of persuading a number of independent authorities to contribute monographs on their specialities. This book is the outcome of such a process, and all Mr. Ratcliffe's team of contributors are among the leaders in their subjects. After an introductory chapter by Sydney Chapman, M. Nicolet discusses the properties and constitution of the upper atmosphere, H. E. Newell describes the results obtained from rockets and satellites, and H. Friedman provides a comprehensive survey of the sun's ionizing radiations. Then follow three linked chapters on the aurora and airglow by D. R. Bates, and a fourth, by H. G. Booker, on radar studies of the aurora. A clear account of the ionosphere is given by J. A. Ratcliffe and K. Weekes; E. H. Vestine deals with the interplay between geomagnetism and the upper atmosphere; and the evidence from meteors is surveyed by J. S. Greenhow and A. C. B. Lovell. Finally, there is a chapter of additional material, entitled "Advances during the IGY, 1957–8 ". The book is particularly admirable on matters of detail: numbered references (over 1200 in all) are given at the end of each chapter; there is a full name-index and a subject-index; misprints are few; the many diagrams are clear, uniform and accurate; the general design and layout are excellent.

All the individual monographs are well-written and valuable, and they are also pleasingly consistent in both matter and manner. Opinions will differ on the relative importance of the subjects discussed, but in general a fair balance has been preserved. There is however one illogicality, which the editor could scarcely avoid: some of the chapters, such as " Radar Studies of the Aurora ", or the results from rockets, are concerned with a specific *technique*, while others, such as " The Ionosphere " or " The Airglow ", deal with a specific *subject*. Consequently, more than one chapter often has to be consulted in order to obtain a full picture of a subject.

The book must be regarded as of 1958 vintage, since most of the chapters were written then. An attempt has been made to bring them up to date by means of supplements in the final chapter, but this procedure is inevitably unsatisfying when major changes are needed, and, furthermore, only five of the eleven contributions carry these postscripts. Also, the editor's apparent claim that the references are complete up to December 1959 cannot be sustained. One suspects too that Chapter 3 was at first intended to be a survey of results from rockets only, the words " and satellites " then being added because events demanded it. In fig. 6 of Chapter 3, for example, values of air density obtained from rockets at heights between 100 and 200 km are unjustifiably extrapolated to 800 km, and the (correct) results from satellites are ' compared ' with these completely erroneous values. Chapter 2 however gives a more up to date picture on this question. There are a few other inconsistencies between chapters. The most blatant concerns the first two figures in the book, on pages 3 and 18, which both deal with atmospheric nomenclature. Such repetition would be less disconcerting if the two diagrams agreed. But they don't: for example, the tropopause is shown at a height of 10 km in the first diagram; in the second it extends from 7 to 17 km, though the 17 could be a misprint for 12. Such minor blemishes do not seriously mar the book's worth: it is the best advanced text on the subject and though its price will daunt the private purchaser, it should find a place in every relevant library.

D. G. KING-HELE.

High Energy Electron Scattering Tables. By ROBERT HERWAN and ROBERT HOFSTADTER. (London: Stanford University Press, Oxford University Press, 1960.) [Pp. 273.] 68*s*.

The bulk of the book consists of a collection of tables of nuclear charge densities, nuclear form factors and scattering cross-sections for the scattering of electrons by nucleons and light nuclei. This should be useful to workers in the field. The introductory section of some 70 pages on electron scattering by nuclei is of more general interest, but is largely extracted from Hofstadter's excellent review articles on the subject.

The book is beautifully produced and, considering its limited appeal, remarkably cheap.

L. R. B. ELTON.

Radio Research 1959. (Published for D.S.I.R. by H.M.S.O., September 1960.) 3*s*.

The advantages to research in radio propagation afforded by space vehicles are referred to in the annual report of the D.S.I.R.'s Radio Research Station. The report of the Director (Dr. R. L. Smith-Rose, C.B.E.) contains details of the Station's work in 1959, including:

Predictions of satellites.

The Station has improved its facilities for positional measurements on satellites, and has extended its service of providing predictions of their orbits. Over one hundred establishments and observers receive these predictions at weekly intervals. Observations on the transmissions from satellites are also being used in studies of the ionosphere. The World Data Centre for rockets and satellites established at the Station in October 1958 is fulfilling its intended purpose of collecting data on space research and exchanging it with the two other centres in Washington and Moscow.

Structure of the ionosphere at high latitudes.

An important part of the Station's investigations following the International Geophysical Year has been on ionospheric phenomena at high latitudes associated with the incidence on the earth's atmosphere of charged particles from the sun, which can cause auroral and magnetic activity, as well as ionospheric disturbances.

Studies have shown that the particle streams move considerable distances towards lower latitudes when the magnetic activity is high, and to higher latitudes when it is low.

International comparison of power-measuring techniques at micro-wave frequencies.

The accurate measurement of power is of fundamental importance at micro-wave frequencies, since in this region of the radio spectrum, the basic quantities, voltage and current, cannot be defined and measured in a satisfactory manner. This is a problem of world-wide interest and the International Scientific Radio Union (URSI) proposed that the Radio Research Station should be the co-ordinating centre for the international comparison of measurement of power at centimetre wavelengths.

High-intensity electromagnetic radiation.

Another field of study is the possible harmful biological effects of high-intensity electromagnetic radiation above about 50 Mc/s. Instruments are being developed to measure the flux density in the neighbourhood of high-power transmitters; the Station's particular work in this field is the provision of known flux densities for the calibration of monitoring instruments.

Refractive properties of the troposphere.

Although long-distance tropospheric scatter propagation is finding increasing application for point-to-point communication the physical processes involved are not yet well understood. A better understanding would enable the planning and engineering of such systems to be improved, and also, facilitate an extension of their use. To this end, and also to increase knowledge of other modes of propagation through the troposphere, the Station has been seeking more detailed information concerning the refractive index structure of the lower atmosphere and factors which influence it. The problem is being tackled in several ways:

(*a*) by an airborne microwave refractometer which records the refractive index variations at heights of up to 3000 metres;

(*b*) by regular 24-hour radar observations every week during the past year and

(*c*) by measurements (made by the Meteorological Office) with a balloon-borne wired sonde.

Energy for the World's Needs

by L. ROTHERHAM
Member for Research, Central Electricity Generating Board

1. INTRODUCTION

The industrial revolution which started in Britain in the 18th century caused major changes in the life of the nation. In Queen Anne's day the main industry was agriculture, but the development of the steam engine, which suddenly put at man's disposal more power than he had ever dreamed of, led to the creation of workshops and factories, etc., and brought about a rapid increase in the population and wealth of the country. Industrialization spread to other countries and today it is still spreading.

Population increase can only take place if there is sufficient food; consequently agriculture has to expand and the farmer must have more power to increase the yield of his land. The increased population will also consume more raw materials. Some industrial countries find it necessary to import food or raw materials from other countries. Thus world trade has developed and the rich countries have been at an advantage over backward or underdeveloped countries in securing favourable balances of trade. For example, today the U.S.A. has to import large quantities of iron-ore from Canada since her own economically workable deposits have been largely used up. From Canadian ore she is able to make earth-moving eqpuiment, drilling rigs, transport, etc., all of which can be sold to Canada and elsewhere to enable more minerals to be extracted from the earth and so on.

The underdeveloped countries find themselves in a difficult situation when they try to improve their economy. Some have little mineral wealth, others have only small energy reserves. Others need only the capital to exploit their resources or to invest in industry; in many such cases development is being assisted by large grants from the more prosperous countries. The present-day pattern of life which will probably continue is one of expanding capital investment, continuing industrialization, and increased mechanization of agriculture to support the growing world population.

2. FORECASTS FOR THE FUTURE

In making forecasts for the future one has to draw on one's knowledge of the past, making some simplifying assumptions. Firstly one has to consider world population and population growth, and world food production.

2.1. *World population*

It is believed that the world population about 10 000 B.C. was under one million, and that the rate of growth was slow up to the time of the Emperor Augustus. Following the decline of the Roman Empire and during the dark ages up to about A.D. 700, it was more or less stationary, probably due to warfare and disease. After A.D. 1400 a slow growth seems to have gathered momentum and the population suddenly expanded in the 17th century in both China and the western hemisphere. The rate of growth (fig. 1) has been regular since then to the present day and the world population is now estimated to be over 2400 million. Growth rates were 0·1 per cent in 1500 and 0·86 per cent in 1950.

If we look at the course of events in Britain (fig. 2) we see the population trend and two isolated events in history: the Roman evacuation of Britain and the black death. At present the population is increasing slowly.

The nearest forecast is that in 100 years time, the world population will have trebled. Such an increase can only occur if there is a major advance in agricultural technology. If the average productivity of the world's agricultural land could be raised to that with the present highest productivity, there is no doubt that a population three times the present size could be fed at least up to current standards. In many parts of the world the rainfall is adequate to supply the drinking and agricultural requirements of a considerably increased population.

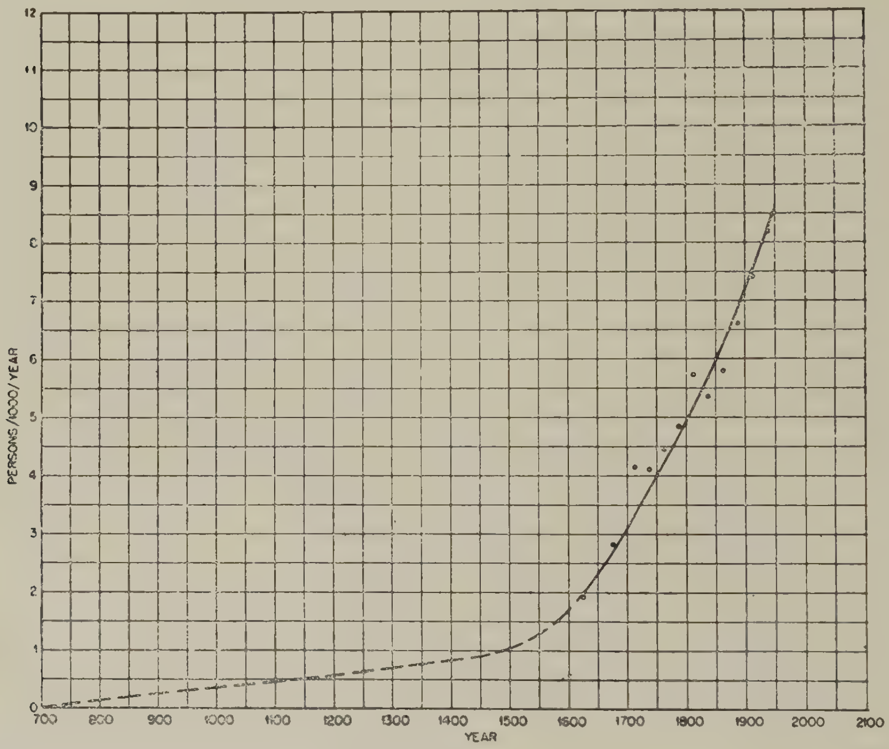


Fig. 1. Rate of growth of world population.

It is estimated that three-quarters of the world's working population are now employed in agriculture and fishing, but the predicted trend towards more industrialization visualizes that in 50 years' time, only half the working population will be so engaged and half will be working in industry and supporting services. This indicates that the present urban development will have doubled in 50 years. This is not impossible since it happened in 65 years in the U.S.A. between 1815 and 1880.

2.2. *Capital expenditure*

Mechanization and an expanding industrial programme can only be accomplished by heavy capital investment. Of the national turnover in Great Britain, approximately 20 per cent is re-invested, of which industry ploughs back less than 10 per cent and thus the rate of expansion is approximately 2 per cent per annum. Russia on the other hand is re-investing up to 40 per cent of her annual turnover, and the results in terms of expansion of oil and steel production and heavy engineering are well known. Other countries, particularly in the east, may take an example from Russia and although their capital resourses are low at the present moment, one might expect their expansion to be more rapid than the average world rate of 1–2 per cent.

In India the planned rate of re-investment is less than 7 per cent at the present time, but it is hoped to increase this to 20 per cent by 1968.

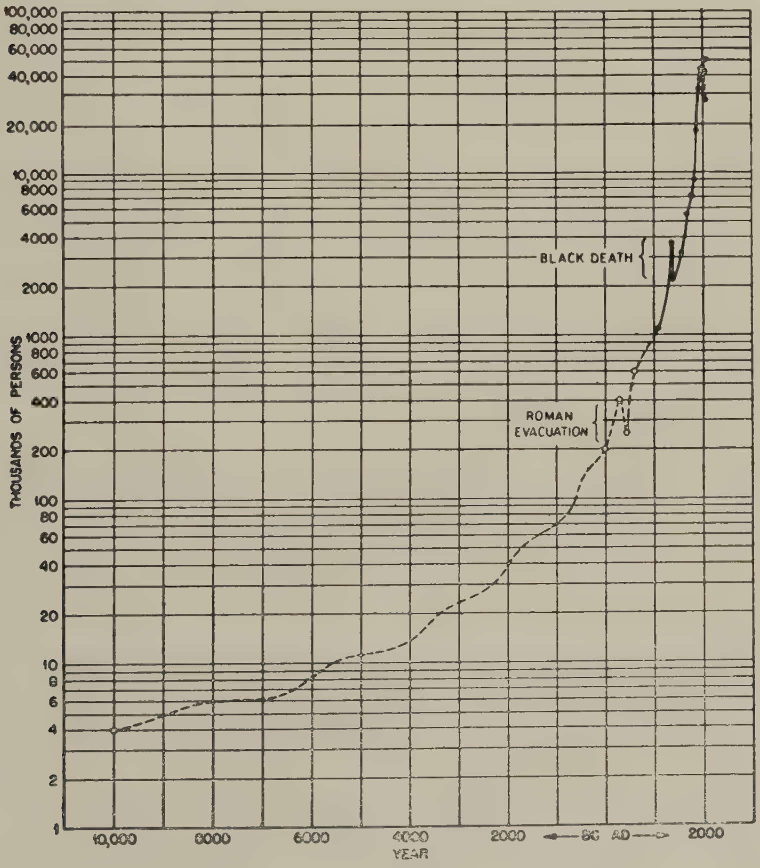


Fig. 2. Population of England and Wales.

Expansion will also be aided by the large loans which are at present being made available to the underdeveloped countries by the United Nations and in particular by the U.S.A.

If the world population expands at the rate indicated above, the capital investment that will be necessary up to the end of the century has been estimated to be £200 000 million of which approximately five-sixths will be for industrialization. The industrial labour force would average about 450 million, therefore the investment per head would be about £350.

A further significant factor is the amount of revenue which various countries allocate to national defence. Thus the United States is spending 64 per cent of her budget revenue on defence, the United Kingdom 24 per cent and India 18 per cent. If it were known that at any time in the future any or all of this money would be invested in industry or to augment a national power programme, our present predictions would have to be drastically revised.

2.3. *Demand for energy*

During the last 100 years, the world demand for energy has grown at a rate between 3 per cent and 4 per cent per annum. However, over the last 50 years this has not meant a proportionate increase in the amount of fuel consumed, because the efficiency of the conversion process has been improved very considerably. For example, in Britain the electricity supply industry has saved 36 million tons of coal in the past 10 years by increasing the average annual thermal efficiency of generation from 21 per cent to approximately 26 per cent. It is expected that stations under construction like those planned will generate at a thermal efficiency of 35 per cent and over. The average world efficiency of electrical generation is estimated to be 22 per cent at the present time. The increases in efficiency have resulted mainly from the use of higher steam pressures and temperatures and by increasing the size of generating sets.

Over the last 50 years approximately two-thirds of the increase in demand for energy has been met by using energy sources more efficiently but we cannot expect the same increases in efficiency to take place indefinitely in the future. The theoretical steam cycle efficiencies on which most electrical generating plants operate are limited to approximately 50 per cent and when allowance is made for the mechanical efficiency of plant it is difficult to see how an overall efficiency of say 45 per cent can be greatly exceeded. The best estimate of the world average efficiency of electrical generation in 100 years' time is approximately 36 per cent, with countries like Britain and America achieving say 43 per cent.

It follows that the world will have to utilize a progressively increasing proportion of the available reserves of energy. Forecasts for the future rates of use are shown in fig. 3 for three different rates of growth of the population. This shows that the world to date has consumed approximately 13×10^{18} B.Th.U. By the end of the next 100 years we will probably have consumed up to 100–400×10^{18} B.Th.U., i.e. the minimum requirement for the next 100 years would appear to be about 87×10^{18} B.Th.U. This is equivalent to more than 11 million million tons of coal burnt at the present efficiency. To put the matter into perspective, half of all the coal ever consumed in the United States has been used in the last 30 years and half of all the oil and gas ever consumed in the United States has been used in the last 15 years. The world in the last 100 years is believed to have consumed one half of all the energy consumed in the preceding 18½ centuries.

If we assume that the increase in the world demand for energy will average 3 per cent per annum over the next 100 years we find that the annual energy consumption by the middle of the 21st century will be approximately 30 times the present amount. If we divide this figure by the estimated population we find that the per capita utilization of energy throughout the world will be approximately twice the present day per capita utilization in the United States. This is tantamount to saying that we can expect in 100 years time that the world will support three times the number of people with an average standard of living twice that of the U.S.A. today.

3. Available energy resources

The forms of energy available comprise two main groups which are generally termed energy capital and energy income.

Examples of energy capital are the fossil fuels, e.g. coal and oil, which are the legacy of the past. Examples of energy income are water power, wind power, animal wastes, solar heat, tidal power and natural steam.

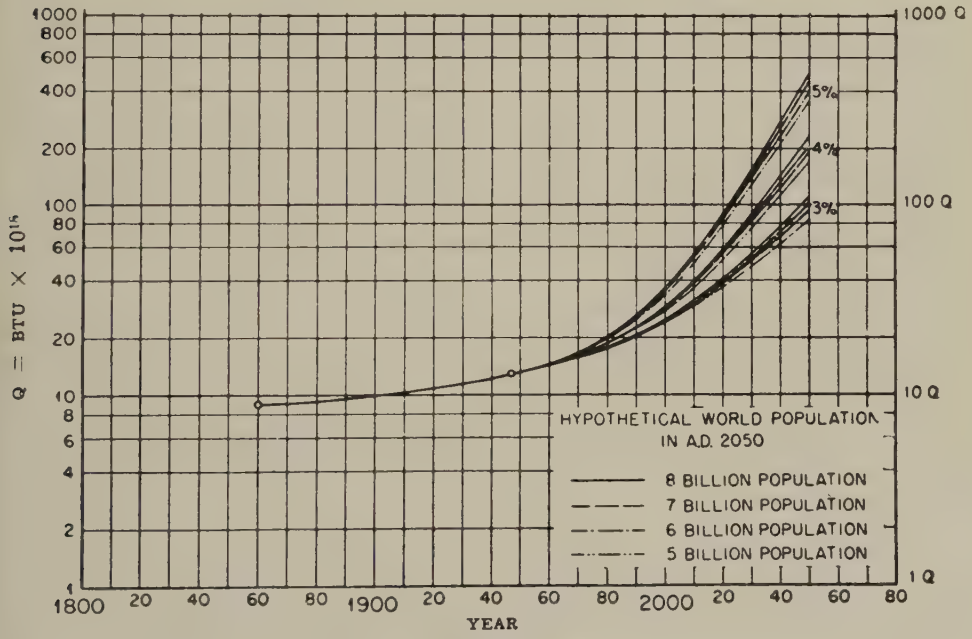


Fig. 3. Cumulative energy requirements of the world.

3.1. *Energy capital*

Under this general heading we can include the various types of coal (anthracite, bituminous coals, brown coals and lignites), petroleum and natural gas, oil shale and peat.

Coals and lignites: Coals are used in the steel industry, and as a raw material in the chemical industry, and thus it is not possible to regard all the available reserves as energy sources. However, the vast majority of coals and lignites are burnt either for space heating or for raising steam in industrial or power station boilers. Approximately 50 per cent of the world's energy requirements are met by coal. Coal production in the United Kingdom at the beginning of this century was approximately 270 million tons per annum, this has now declined to approximately 200 million tons per annum. Of this total, some 40 million tons per annum are used by the Central Electricity Generating Board. The total estimated reserves of coals and lignites for three countries and for the world are given below:

Country	Coal resources tons × 10^9
U.S.A.	1720
United Kingdom	170
India	62
World total	4040

It has been estimated that only one-third to one-quarter of the coal reserves could be economically brought to the surface (economically is defined by Putnam as costing not more than twice present-day costs). At the present rate of world consumption, the workable reserve is estimated to last approximately 500 years —that in the United Kingdom at the present rate of consumption between 100 and 250 years.

Petroleum and natural gas: At present oil supplies one-quarter of the world's energy requirements and natural gas about one-ninth. The demand for oil is growing rapidly, largely owing to the prominence of the internal combustion engine and the gas turbine.

New reserves of oil have been discovered in greater quantities as surveys have advanced, and the estimates of the remaining resources have steadily increased. At the beginning of 1956, remaining proved reserves were estimated to be about 26 000 million tons; at the end of 1958 the estimate had risen to over 37 000 million tons.

From past experience it seems reasonable to assume that the thermal value of the natural gas in oil bearing strata is about 60 per cent of that of the oil. Reserves of methane associated with coal have been estimated to be equivalent in thermal value to about 1 per cent of that of the coal.

Oil shale: In many parts of the world there are large deposits of oil shale and similar materials, including bituminous tar sands, which on heating or extraction yield oil. Estimates of the quantity of oil that might be recovered by conventional methods from such materials have ranged from 20 000 million tons to more than 40 000 million tons. Of these estimated reserves more than one-half are in the U.S.A. and Canada, and one-third in the U.S.S.R. Owing to the high cost of production in comparison with petroleum, the amount of oil obtained from shale has not so far been more than about 300 000 tons a year, of which one-third has been produced in Scotland.

Experiments are planned or in progress to recover oil using underground nuclear explosions and by such methods as pumping hot water and detergents to loosen the oil from the rock, injecting air and igniting part of the oil, gas drive and thermal methods, all of which should considerably raise the estimate of recoverable reserves.

3.2. *Energy income*

Water power: The world water power potential excluding tidal power has been estimated to be approximately 1500 million kilowatts. To generate this quantity electrically at the present average efficiency of thermal power stations would require approximately 3000 million tons of coal a year assuming a load factor of 50 per cent. The present installed capacity throughout the world is 100 million kilowatts corresponding to an annual consumption of 230 million tons of coal or 150 million tons of fuel oil. Further hydro-electric schemes are under construction where the costs are reasonable, but because the easily developed sites have largely been utilized and the remaining sites tend to be remote from the load centres and/or in difficult mountainous country, the capital costs of hydro-electric stations and transmission lines will tend to increase.

In Britain the only suitable sites for hydro-electric stations are in Scotland, and to a smaller extent in Wales. Although the output from hydro stations in Britain provides less than 2 per cent of the total electrical energy requirements, compared with 99 per cent in Norway, 20 per cent in the U.S.A. and 5 per cent in India, the schemes in North and Central Scotland have been of considerable importance to the local economy. During the last 12 years, 25 power stations with installed plant of approximately 600 000 kilowatts have been commissioned.

It has been estimated that the power corresponding to the dissipation of the tides is 1000 million kilowatts. If this power could be economically harnessed it would be equivalent to burning 2000 million tons of coal a year in thermal power stations, i.e. it would just about meet the present day world demand for coal. Several tidal power schemes have been considered in various parts of the world where the tidal range is high. In England the Severn barrage has been considered several times in the last 30 years but the only scheme on which construction has begun is in the Rance estuary near St. Malo, France. Here one turbine is working experimentally but, because of the discovery of natural gas deposits in the Pyrenees, the whole scheme is being reconsidered. If the full plans are implemented it is estimated that 800 million kilowatt hours of electricity will be produced per annum.

Wind power: Estimates of wind power available on the earth's surface are extremely difficult to make but it is believed that there is about three times as much power in the wind as in the tides although a much lower proportion of wind power is likely to be usefully applied. Windmills have been used for centuries to produce small amounts of power, but many of them have fallen into disuse within the last 100 years. However, during recent years the possibilities of economic utilization of wind power have been further investigated in several regions (fig. 4).

In this country active interest in a wind generator led to the development of a 100 kilowatt machine in the early 1950's. An extensive series of tests of a wind generator were carried out in the Orkney Isles and it operated reasonably satisfactorily. However, the winds around our coasts are irregular, and the capital cost of the installation was high compared with other forms of electrical generation. As a result the machine was later transferred to Algeria where it is now operating regularly at twice its design output.

Vegetable and animal wastes: There are numerous vegetable and animal wastes which are used to varying extents as fuel for the generation of heat and power. Examples are sawdust, wood shavings, straw, sewage sludge, town refuse and animal dung. In South Africa, boilers as large as those used in some of our power stations use wood wastes for fuel in paper producing plants. Pass-out turbines supply steam to heat the wood pulp, and the electricity generated supplies the motors which drive the saws and rolling machines.

In South America, furnaces have been designed to burn wood, sugar cane waste, coffee husks, rice husks and straw. To anyone unaccustomed to think of these materials as fuels, it might seem that the processes are inherently inefficient, but their calorific values are of the order of 80 per cent of that of an average coal.

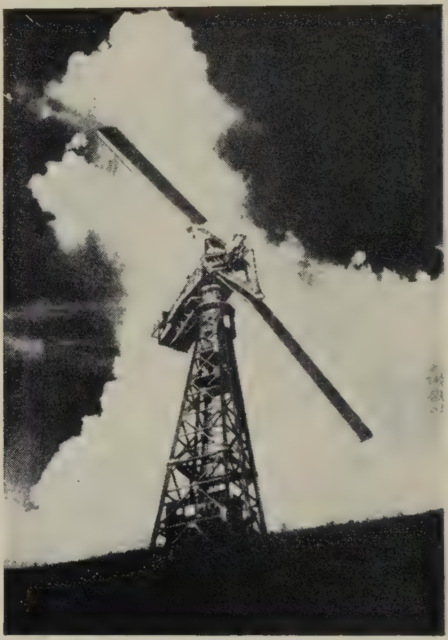


Fig. 4. A 1250 kw wind generator in the U.S.A.

Figure 5 shows the energy at present derived from farm wastes compared with that from coal, oil, water, wood and electricity.

Methane is a gaseous fuel that can be obtained by the fermentation of vegetable and animal matter under controlled conditions. The gas is being collected and utilized at several sewage disposal works in Great Britain and other countries during the fermentation of sewage sludge. In Great Britain it has a total annual thermal value equivalent to approximately 25 000 tons of coal.

In France, methane from farmyard manure is used for domestic cooking, heating and lighting. Experiments in Germany and East Africa indicate that sisal waste may be useful for methane production: 10 tons of sisal waste could yield gas equivalent to 800 gallons of oil.

In many parts of the world animal waste is used directly as a source of heat. Millions of people in India and the other Asian countries, Central and North Africa and Egypt know no other way of cooking their meals. However, the use of such materials as fuel is uneconomic compared with their value as fertilizers and increasingly they will have to be put back into the land. The maximum amount of heat that will be obtained from animal wastes in the future does not appear likely therefore to exceed the equivalent of 500 million tons of coal per annum.

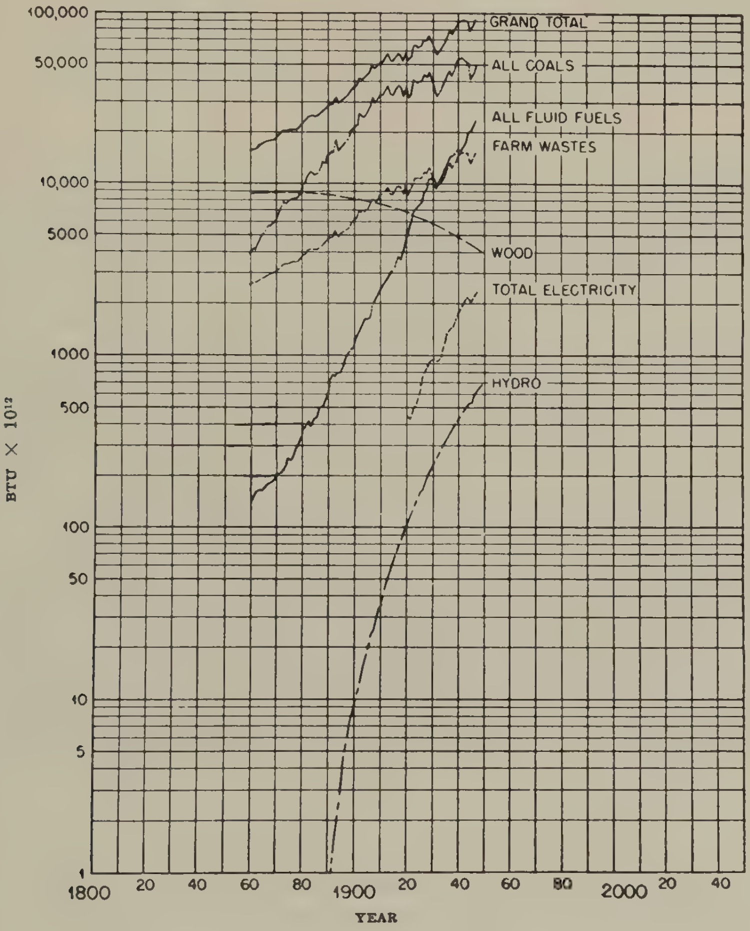


Fig. 5. Energy derived from different sources.

Solar heat: Basically most forms of energy are derived from the sun, which is radiating energy at a rate equivalent to burning more than 10 000 million million tons of coal every second. Of this energy only one part in 2000 million reaches the earth's atmosphere and quite a large proportion of this amount is reflected back into space. Even so, the energy received annually by the surface of the earth from the sun is approximately 5000 times the heat of combustion of all the solid liquid, and gaseous fuels used annually in the world at the present time.

The problem is to concentrate the radiation into a large unit source of power. This has been attempted by using mirrors or reflectors to concentrate the sun's rays. Electricity has been generated by these means using the concentrated rays to heat water and generate steam. Some countries already use solar energy for commercial purposes and many tropical countries use it domestically. The obvious limitation of this form of energy is in temperate climes where clouds limit the radiation received.

Natural steam: Geothermal heat in the form of natural steam is available in Iceland, Mexico, New Zealand, Japan and Italy. The steam is often heavily contaminated with minerals which make it highly corrosive, so it cannot always be used directly to drive steam turbines. Nevertheless, Italy has 400 000 kilowatts of plant installed in power stations utilizing natural steam. In Iceland, where power is generated almost entirely by hydro-electric means, a high proportion of the houses in Reykjavik are heated by geothermal energy. In New Zealand a scheme to utilize natural steam to generate electricity is under way. Plant to generate 40–50 Mw is already working and additional sets are now being commissioned: a total capacity of 150 Mw has so far been sanctioned.

3.3. *Nuclear energy*

The various energy resources considered above are summarized in fig. 5, but one of the major questions is how much of this energy is economically recoverable. If one assumes that the price of energy over the next 100 years must not rise to more than twice the present-day value, it appears that little more than one million million tons of coal are economically available as capital energy. If the forecasts of world population given above are correct, and likewise the forecasts for the increase in the demand for energy are correct, then over the next 100 years at best we would only be able to meet 16 per cent of our energy requirements from income energy. However, at the present time energy represents less than 2 per cent of the total national income and it may be that we can afford to let the cost rise to more than twice the current value. This would mean an increase in the economically recoverable resources of capital energy.

Therefore recognizing the fact that in 100 years' time, the world is likely to be consuming energy at nearly 30 times the present-day rate, a continuing increase in the standard of living cannot take place unless the conventional energy resources are supplemented to a continuingly greater degree by some new form of energy.

The only logical way in which the expanding demand for power can be met, would appear to be a steady increase in the amount of electrical energy which can be made available. Electricity is clean and convenient for use both in industry and the home. The cost of electricity has remained relatively stable compared with the increase in the cost of coal and the general increase in the cost of living.

The only form of energy not so far discussed, which fulfils the criteria of no great increase in cost and extensive resources, is nuclear energy, and this can be harnessed and converted into electrical power.

Nuclear energy results from the forces which maintain atomic particles in the nucleus of an atom. There are two main energy releasing processes, namely, fission and fusion. In the fission process, atoms of particular elements are split by bombarding them with neutrons and in splitting they release energy, e.g. 1 lb of the 235 isotope of uranium (which is the only naturally occurring material which is spontaneously fissionable) releases as much energy as 1400 tons of coal.

The fusion process consists of bringing together atoms of similar light elements which under certain conditions form a separate heavier element and release energy. The conversion in this way of hydrogen into helium is believed to account for the energy emitted by the sun and some stars. The energy released in forming 1 lb of helium from hydrogen is equivalent to 10 400 tons of coal.

Harnessing the fusion process for the supply of useful energy has not yet been achieved but the fission process is already being applied in Great Britain and other countries to produce electricity (fig. 6).

Fig. 6. Impression of Bradwell nuclear power station, with six 52 Mw sets.

Nuclear energy reserves: The fuels at present being used in nuclear power reactors are uranium and thorium. Although uranium is a common element widely distributed in the earth's crust, concentrated deposits tend to be rare, but those containing only a fraction of 1 per cent of uranium are considered to be economically exploitable. It has already been established that adequate supplies exist far in excess of immediate demands and the tempo of prospecting has therefore declined. Outside Communist countries world production of uranium oxide (U^{238}) has risen from 11 000 tons in 1955 to nearly 40 000 tons at present. Principal reserves of U^{238} outside Communist countries are estimated to be about 1 million tons and are mainly found in Canada, South Africa, U.S.A., France, Australia and the Belgian Congo.

Corresponding reserves of thorium as ThO_2 are estimated to be about 850 000 tons and deposits exist in India, Canada, Australia, U.S.A., Brazil, Nyasaland and South and West Africa.

Earlier in this paper energy resources were divided into two classes; capital and income. It is interesting to ask into which category uranium should be placed. At first glance it would seem reasonable to call it a capital resource because in time it will be used up. However, development in the design of nuclear reactors is leading to advanced types, the breeder reactors, which, it is hoped, will produce more fuel than they burn. Thus, uranium used in a fast breeder reactor will be partially converted into plutonium which itself is a nuclear fuel. This bonus of fuel might well be regarded as an energy income.

As in the case of coal and oil, nuclear fuels have alternative uses as raw materials for the chemical and other process industries, so all the usable reserves must not be regarded as energy sources.

The very high specific energy of the nuclear fuels confers an advantage in that the fuels are easily transportable. Water and wind power must be used *in situ* and there is a limit to the economical transmission of electricity over large distances; coal, oil and gas are expensive to transport. Nuclear energy resources thus not only compose a significant proportion of the world's reserves, but offer the best means of bringing power to underdeveloped countries providing that adequate capital is forthcoming.

We can say little about the fusion process at the moment but if it becomes practicable the reserves of hydrogen to sustain it are almost limitless.

4. Conclusions

The figures quoted are of course mainly estimates which are bound to contain considerable inaccuracies, but allowing for this and assuming that no unforeseen catastrophe will occur, the general picture is probably not far wrong.

World population is still increasing and nations are endeavouring to improve their living standards. At present rates of economic and population growth, about 30 times the present annual energy requirements will be needed in 100 years' time to support a trebled population with a very high average standard of living.

The power can be provided from known energy resources but the main problem will be to harness them and to do so economically.

What of the years that will follow the 21st century? There comes a point where extrapolation of present trends leads to gross inaccuracies and meaningless forecasts but we can perhaps take an optimistic view. In the very recent past we have seen the fears of a shortage of uranium dissipated to such an extent that prospecting for reserves has diminished. The wastes of the Sahara have just begun to yield oil and hundreds of millions of tons lie beneath the desert. Searches for other mineral wealth have revealed the enormous reserves of iron ore in Venezuela. Who knows then what other energy reserves and mineral wealth lie undiscovered beneath the land, sea and ice in the deserted tracts of the earth's surface? Our optimism springs from the fact that every predicted shortage of energy reserves will stimulate speculative prospecting and history shows that, in the aggregate, prospectors always seem to ' strike it rich '.

Faraday and the Structure of Matter

by L. PEARCE WILLIAMS

Department of History,
Cornell University, Ithaca, New York.

SUMMARY

Michael Faraday, the experimentalist, is a well-known and beloved figure in the history of science; Michael Faraday, the theorist, is almost a stranger. The purpose of this paper is to describe Faraday's theoretical ideas on the structure of matter and to illustrate his use of these ideas as guides in his researches on electrolysis, static induction and ordinary electrical conduction.

1. INTRODUCTION

Michael Faraday is rightly regarded as the greatest experimentalist in the history of science. The three volumes of the *Experimental Researches in Electricity* are not only one of the master works of science but a unique record of sustained experimental investigation. And the achievement is awesome; with what patience did Faraday coax and cajole Nature to reveal her secrets; with what skill did he torture each new fact to ensure that it was a fact and not merely an illusion!

But Faraday's very skill as an experimenter, together with his apparent distrust of theory, served to create a myth which still persists. It is the myth that Faraday was the perfect Baconian. His success was attributed to his seeming willingness to remain an objective collector of facts. Since Faraday's career coincided with the revival of Baconianism considered as a (indeed *the*) proper scientific method, the origins of the myth are easily explained. Like most myths, it has endured far beyond the time when it might be supposed to reflect reality. For the creators of the myth, Faraday's greatness *did* lie in his discovery of new facts such as the rotation of a current-carrying wire around a magnetic pole, electromagnetic induction, diamagnetism, and the rotation of the plane of polarized light in a magnetic field. Since the publication of Einstein's General Theory of Relativity, however, we must see Faraday in a different perspective. Can the foundations of field theory be only the accidental by-products of clever laboratory manipulation, or must we not recognize that the hand of the skilful experimenter was directed by the keenest theoretical mind of the nineteenth century? Actually, each discovery had been suggested by theoretical considerations, and each newly discovered fact became, in turn, part of the structure of theory. It is not my purpose here to examine this theoretical structure in its entirety for that would necessitate going far beyond the limits of a short article. What I should like to do is present Faraday's ideas on the constitution and structure of matter and show, in some detail, how these served as the guide for his first important series of experimental investigations.

2. FARADAY AND THE ATOMIC THEORY

Faraday's attitude toward the atomic theory has generally been used to illustrate his caution and distrust of ' hypotheses '. In 1844, when the atomic theory was the acknowledged basis of chemistry, with the full weight of the authority of Berzelius (and most chemists) behind it, Faraday wrote:

> " The view of the atomic constitution of matter which I think is most prevalent, is that which considers the atom as a something material having a certain volume, upon which those powers were impressed at the creation, which have given it, from that time to the present, the capability of constituting, when many atoms are congregated together into groups, the different substances whose effects and properties we observe. These, though grouped and held together by their powers, do not touch each other, but have intervening space, otherwise pressure or cold could not make a body contract into a smaller bulk, nor heat or tension make it larger; in liquids these atoms or particles are free to move about one another, and in vapours or gases they are also present, but removed very much further apart, though still related to each other by their powers.
>
> " The atomic doctrine is greatly used one way or another in this, our day, for the interpretation of phenomena, especially those of crystallography and chemistry, and is not so carefully distinguished from the facts, but that it often appears to him who stands in the position of student, as a statement of the facts themselves, though it is at best but an assumption; of the truth of which we can assert nothing, whatever we may say or think of its probability. The word atom, which can never be used without involving much that is purely hypothetical, is often *intended* to be used to express a simple fact; but good as the intention is, I have not yet found a mind that did habitually separate it from its accompanying temptations; and there can be no doubt that the words definite proportions, equivalents, primes, etc., which did and do express fully all the *facts* of what is usually called the atomic theory in chemistry, were dismissed because they were not expressive enough, and did not say all that was in the mind of him who used the word atom in their stead; they did not express the hypothesis as well as the fact."

Faraday's objection to the atomic theory was not only that it went beyond experience but also that it failed to account for a whole host of phenomena. By 1844, indeed, it was almost impossible to speak of *an* atomic theory. There were different kinds of atoms, each suited to account for certain observed facts, but none sufficient to explain them all. To account for chemical combination in definite proportions, Dalton had introduced his ' billiard ball ' atoms which by their indivisibility, gave an ultimate physical reality to these combining ratios. From his work on gases, he also recognized the necessity of having a repulsive principle to account for the rarefaction and expansibility of aeriform substances; thus, each 'billiard ball', endowed with the attractive principle of universal gravitation, was surrounded by an atmosphere of caloric (the ' fluid ' of heat) which repelled other similarly equipped atoms. The discovery of the electrolytic properties of the voltaic pile led Berzelius to suggest an atom whose opposite ends contained equal charges of positive and negative electricity. The atom, as a whole, was electrically neutral, yet would respond to the electrical forces of the pile and enable one to account for the observable electrochemical decompositions. After Oersted's demonstration of the magnetic effects of an electric current in 1820, Ampère proposed another kind of atom in which a positively charged nucleus was surrounded by a current of negative electricity. Magnetic effects were thereby reduced to the results of the alignment of these electro-dynamic atoms. Faraday's own discovery of diamagnetism gave rise to still another atom in which the electric currents were so arranged that the particles would set themselves across, rather than along, the lines of magnetic force.

This proliferation of atomic models when thus baldly stated would appear to militate against the acceptance of the atomic theory by any thinking person in the nineteenth century. But with the exception of a few men like Faraday, such a confrontation of mutually incompatible schemes rarely took place. In fact, they proved most useful in the limited areas to which each concept was applied by the active research scientist. Not the least valuable of the many lessons of the history of science is that of the importance of ignoring ' details ' which get in one's way. For the analytical chemist, ' billiard balls ' were sufficient and if it was objected that they could not explain the origin of electric or magnetic forces, the tendency was for him to shrug his shoulders and suggest vaguely that an imponderable electric or magnetic fluid could take care of that difficulty. Similarly, the physicist interested in electrical effects could more or less ignore the problems of chemical combination in *his* atomic theory. The fruits of this method should be noted. By disregarding side issues, the analytical chemists produced the periodic table; from the physicist came the discovery of the electron and the ' sub-atomic ' particles.

To Faraday, however, the situation was more complicated. Like Berzelius, he was an analytical chemist of considerable skill; like Gay-Lussac and Dalton, he was applauded by the scientific community for his work on gases; like Oersted and Ampère, he created a new epoch in the study of electromagnetism; like Fresnel and Young, he contributed fundamentally to the theory of light; like Sir Humphry Davy, he was a founder of electrochemistry. Unlike these men, he worked almost simultaneously in all these fields. The annoying but seemingly irrelevant ' detail ' of one area loomed as an impassable barrier in another. As early as 1817, Faraday had recognized the dangerous (and human) desire to theorize over-hastily thereby gaining present certainty at the expense of future error. " Better " he wrote " to own our ignorance of these phenomena, if we cannot by their means ascertain satisfactorily the true circumstances of the case, than to form an hypothesis which shall in accounting for a single instance give birth to a thousand chimera elsewhere."

Faraday, keenly aware of the inadequacies of the commonly held atomic theories, did not suggest (as some did) that all such theoretical considerations be expelled from science. Quite the contrary; part of his hostility to the ordinary concept of atoms arose from the fact that he was the ardent promoter of another atomic theory which, he felt, did account adequately for phenomena. Moreover, this theory had the further advantage, in Faraday's eyes, of being based on a minimum of speculation and hypothesis.

> " If we must assume at all, as indeed in a branch of knowledge like the present we can hardly help it, then the safest course appears to be to assume as little as possible, and in that respect the atoms of Boscovich appear to me to have a great advantage over the more usual notion. His atoms, if I understand aright, are mere centres of forces or powers, not particles of matter in which the powers themselves reside. If, in the ordinary view of atoms, we call the particle of matter away from the powers a, and the system of powers or forces in and around it m, then in Boscovich's theory a disappears, or is a mere mathematical point, whilst in the usual notion it is a little unchangeable, impenetrable piece of matter, and m is an atmosphere of force grouped around it.
>
> " In many of the hypothetical uses made of atoms, as in crystallography, chemistry, magnetism, etc., this difference in the assumption makes little or no alteration in the results, but in other cases, as of electric conduction, the nature of light, the manner in which bodies combine to produce compounds, the effects of forces, as heat or electricity, upon matter, the difference will be very great."

It is my contention that Boscovich's atomic theory provided the basic theoretical framework for Faraday's most important work, at least until the discovery of diamagnetism in 1845. If we are to understand Faraday's development, therefore, we must first examine this theory and the circumstances in which Faraday became acquainted with it.

3. Boscovich's atomic theory

In 1763, Father Roger Joseph Boscovich, S.J., published his *Theoria Naturalis* in Venice. In this work he attempted to avoid some of the problems implicit in the conventional (Newtonian) corpuscular philosophy. The question of the mechanism of impact was of particular urgency, for Boscovich saw that it was not possible merely to reduce billiard balls to atomic size and let them bounce into one another. Ultimate particles must, by definition, be incompressible for there is no space into which the parts (which cannot exist) of an atom can be compressed. What, then, happens when two atoms collide? Without compression and elastic deformation, impact and rebound must be instantaneous. If atom A strikes B, at the moment of impact A will have two velocities, which is absurd. To avoid this absurdity, Boscovich eliminated the material aspect of the atom and reduced it to a mathematical point surrounded by alternating shells of attractive and repulsive forces. In fig. 1, OB is a measure of the distance from the mathematical point-centre of force at O; repulsive forces are represented by the loops of the curve above OB; attractive forces by the loops below OB. At C and H these alternating forces change character. AC is asymptotic to OY. As one approaches O the repulsive force approaches infinity, thus preserving the property of impenetrability. At H, receding from

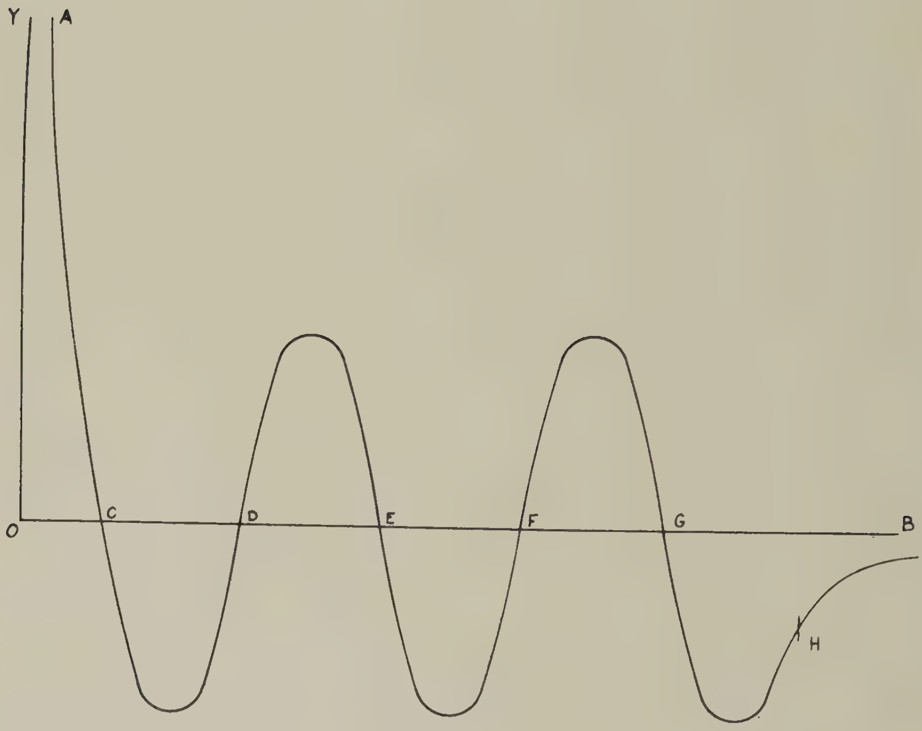


Fig. 1. Boscovich's curve of atomic forces.

the centre of force, the curve is transformed into a hyperbola, asymptotic to OB and representing the inverse square law of universal gravitation. Boscovich carefully avoided putting dimensions on this figure, except to make a basic dividing line at H. To the right of H we are in the macroscopic world; to the left is the realm of atomic interaction.

Boscovichean atoms retained all the properties of the Newtonian corpuscles. The repulsive forces gave them solidity, impenetrability and elasticity; the existence of the forces in space provided extension; the atoms gravitated towards one another and followed the Newtonian laws of motion. More important than the similarities were the differences, for it was these which gave the Boscovichean system its considerable advantage over the older scheme. This advantage was particularly striking in the area of chemistry. Consider, for a moment, the problem of the origin of the perceptible qualities of bodies. This problem had occupied chemical philosophers since antiquity when the atomic hypothesis was first proposed. Two general atomic schemes had been suggested, both ultimately locating the origin of sensible qualities in atomic (or molecular) configuration. Thus, the chemist could assume either differently shaped atoms or round atoms aggregated into definite shapes and then attempt to explain chemical phenomena by means of these geometrical forms. For example, acids were able to corrode other bodies, because they had sharp points and could penetrate the narrow pores of dense bodies such as metals. The difficulty in this system lay in accounting for the regularity of the chemical forms. Why, if universal gravitation were the only force at work, should there be the production of geometrically perfect (and for each compound, constant) forms instead of amorphous clumps? And why should there be ' elective attractions ' so that a substance, inert in the presence of one reagent, reacted vigorously with another? Such questions were unanswerable in the old system and even led Lavoisier to reject the whole atomic speculation as essentially metaphysical and meaningless. With Boscovichean atoms, however, these problems were easily solved. From the primitive centres of force, more complex and geometrically regular assemblages of atoms could be built up whose stability was the result of the complex interaction of the atomic forces. These bodies, the chemical elements, reacted in turn with others to create molecular forms, again rigidly determined by the patterns of the forces involved. Reactions took place when the forces of two elementary particles could mesh in such a way that a stable compound could be formed. The stability of the compound was determined by the strength of the forces involved.

Even on the macroscopic level, the problem of form was more easily handled with Boscovichean atoms than with material particles. How and why did particles assume the regular positions they did in crystalline bodies? If material bodies were affected only by the attractive force of gravitation and the repulsive force of caloric how could one account for the constancy of crystalline form exhibited by mineral species? Once again, the Boscovichean system removed these difficulties by considering geometrical form as the result of the complex interlocking of atomic or molecular forces.

Problems of change of state, too, could be more easily understood in the Boscovichean system. The stability of physical states over a relatively wide range of temperature could be directly related to the alternating attractive and repulsive forces of the constituent particles. The specific and constant transition points (i.e. melting and boiling points) could be located as the various neutral points on the Boscovichean curve.

When Boscovich presented his ideas, chemists were in no position to utilize them. As the eighteenth century advanced, the general current of chemical thought ran increasingly stronger against this kind of ' metaphysical ' speculation, which, though seductive, appeared to be beyond the realm of experimental verification. As Lavoisier wrote in his *Elementary Treatise on Chemistry*,

> "All that can be said about the number and nature of elements seems to me to be discussions which are purely metaphysical, to be indeterminate problems capable of an infinite number of solutions, not one of which, in all probability is consistent with nature. I shall, therefore, content myself with saying that if, by the term elements we mean to express the simple and indivisible molecules that compose bodies, it is probable that we know nothing about them. . . ."

The firm philosophical foundation on which Lavoisier wished to place chemical science derived from this attitude; chemistry was a science of *facts*, not theories. Elements were the last term of a laboratory analysis; ' principles ' were not metaphysical entities (like pointed acid particles) but ponderable substances whose presence conferred certain properties upon compounds containing them (i.e. oxygen as the acid-forming substance presumed to exist in all acids). In this scheme there was no room for Boscovich.

When Lavoisier's system (but not his method) came tumbling down with the discovery of the elementary nature of chlorine and, hence, of the non-existence of oxygen in one of the strongest acids, the old problem of the origin of chemical properties once more became intrusive. Furthermore, with the discovery of the voltaic pile in 1799 and the realization that electricity played a major role in chemical combinations, there was a strongly-felt desire for some theoretical background against which the host of new discoveries which followed could be viewed. It is, I think, no coincidence that Sir Humphry Davy, who established chlorine as an element and whose work in electrochemistry was fundamental was a strong Boscovichean. Just when he became a proponent of Boscovichean atomism I have not been able to determine, but his first serious use of it was after he became convinced that chlorine was not a compound of some unknown base and oxygen as demanded by Lavoisier's theory. Now, if chlorine were an element, then chemistry was in the rather uncomfortable situation of being unable even to give a definition of an acid. The best one could do was say that an acid was a substance which turned litmus pink. Davy, rather tentatively at first, and then with more and more confidence, insisted that acidity was the result of molecular form rather than of the presence of some ponderable or imponderable ' acidifier '. With this in mind, he turned to the analysis of the acids of phosphorus, but without notable results. With more success, he established that the diamond was pure carbon. The enormous differences in the properties of the diamond and ordinary carbon were solely the result of the modification of the geometrical arrangement of the carbon particles.

During these years, Faraday was Davy's assistant, working at the master's elbow and eagerly snatching up everything he could in the way of chemical theory and practice. There can be little doubt that it was through Davy that Faraday was first introduced to Boscovich's theory. In Davy's *Common-Place Book* for 1815–16, in which both Faraday and Davy wrote notes and jottings, there is a passage, probably by Faraday, in which Boscovich's theory is used to explain ' all the phenomena of chemistry '. In a series of lectures on chemistry which he delivered before the City Philosophical Society, a group of young men dedicated to self-improvement to which he belonged, Faraday again brought forward the doctrine of immaterial atoms as being worthy of consideration by his audience. For the most part, however, during the years from 1816, when he published his first paper, to 1831 when he began his classical researches on electrochemistry, Faraday was occupied with matters which strained his practical abilities to the utmost but required little in the way of theoretical reasonings. We must, of course, except his discovery of the rotation of a current-carrying wire around a magnetic pole in 1821 and his discovery of electromagnetic induction in 1831. Even here, the role of theory was rather limited. Faraday himself tells us that it was while writing a history of electromagnetism that he saw that a current-carrying wire should rotate around a magnetic pole and, with his unusual manipulative skill and insight, envisioned the apparatus necessary to illustrate the expected effect. The discovery of electromagnetic induction derived almost entirely from Faraday's deeply felt conviction of the symmetry of nature; if, as Oersted had shown, an electric current could produce magnetic effects then the converse must be true and the real problem was to devise the proper apparatus to make the effect manifest. For the rest, this period was almost entirely devoted to running rather standard qualitative or quantitative analyses and to the highly empirical investigations on the alloys of steel and on glass manufacture.

There is one discovery during this time that deserves special mention for it is the first instance, to my knowledge, where the theory of immaterial atoms led to a specific and new result. It was Davy, however, and not Faraday who foresaw the unexpected. Faraday had been working for some time with crystals of the hydrate of chlorine when, in the early spring of 1823, Davy suggested that he heat the crystals in a closed glass tube. The result was the liquefaction of chlorine which both Davy and Faraday mistakenly thought to be the first condensation of a so-called ' permanent ' gas. Davy, in a note appended to the paper announcing this discovery, wrote:

> " In desiring Mr. Faraday to expose the hydrate of chlorine to heat in a closed glass tube, it occurred to me that one of three things would happen: that it would become fluid as a hydrate; or that a decomposition of water would occur, and euchlorine and muriatic acid be formed; or that the chlorine would separate in a condensed state."

If we believe Davy, and in the absence of any evidence for disbelief I think we must, then we have to ask why he felt that the application of heat could lead to the liquefaction of chlorine. According to the Daltonian theory such a result would be quite impossible, for the addition of heat, the principle of repulsion, should lead rather to rarefaction than condensation. Davy's suggestion, however, makes rather good sense when looked at from the viewpoint of immaterial centres of force. The chlorine particles, bound in the crystals of the hydrate, existed in close proximity to one another and to the particles of water. To produce liquid chlorine, the water must be driven off and the chlorine particles stabilized at the atomic distance of a liquid. Davy, holding heat to be the motion of the constituent particles of a substance, no doubt felt that the slow increase of molecular agitation might loosen the bonds by which the water particles were held and force the chlorine molecules into another, more stable and mutually more intimate association. Referring to the curve in fig. 1, heat as motion might push the chlorine particles over the ' hump ' of repulsion, from say G to F, thus reducing the whole to a liquid.

This use of the theory of immaterial atoms should be noted for there can be no doubt that Faraday's acceptance of it came about largely as a result of witnessing its utility and power in Davy's hands. There were three aspects of the theory which appealed irresistibly to Faraday. There was the fact that, although hypothetical, it was less hypothetical than competing theories. Faraday felt, as mentioned above " The safest course appears to be to assume as little as possible " and immaterial atoms came closest to reflecting reality:

> " . . . all our perception and knowledge of the atom, and even our fancy, is limited to ideas of its powers: what thought remains on which to hang the imagination of (a nucleus) independent of the acknowledged forces? A mind just entering on the subject may consider it difficult to think of the powers of matter independent of a separate something to be called *the matter*, but it is certainly far more difficult, and indeed impossible, to think of or imagine that *matter*, independent of the powers. Now the powers we know and recognize in every phenomenon of the creation, the abstract matter in none; why then assume the existence of that of which we are ignorant, which we cannot conceive, and for which there is no philosophical necessity? "

The theory recommended itself also by the fact that it not only accounted for phenomena in a more complete manner than any other general hypothesis, but that it also provided a basic point of view from which to see physical reality. Davy was one of the first in the nineteenth century to realize the importance of structure in explaining sensible qualities. As we shall see, when Faraday discovered some new effect, his immediate thought was that it involved either a peculiar form or change in form of the arrangement of the constituent particles.

Finally there was the ability of the theory to predict new effects and, as it actually led to new discoveries this reinforced its claims to truth in Faraday's mind. The hold this hypothesis had on Faraday cannot be over-emphasized. He thought, almost instinctively, in terms of immaterial atoms and there are many passages in his works which make little or no sense unless one is aware of Faraday's theoretical frame of reference.

4. ELECTROLYSIS, STATIC INDUCTION AND CONDUCTION

In order to illustrate how Faraday used the Boscovichean concept, let us now examine one phase of the mighty achievement set down in the *Experimental Researches in Electricity*. Since Faraday carefully built his theory, each new fact being added in its proper place to the edifice of theory, it would be well to select the illustrative example from the earlier researches. Let us, therefore, consider the series of researches in the 1830's in which Faraday revealed the basic unity of electrolytic conduction, electrostatic induction and ordinary electrical conduction, and in which, by challenging the orthodox view of action at a distance, he laid the foundations for the modern theory of the field.

In order properly to understand the daring and originality of Faraday, we must first briefly glance at the conventional ideas of these three aspects of electrical science. The almost universal explanation of all electrical effects before Faraday was to be found in the assumption of one or two imponderable electrical fluids whose attractions and repulsions, under specified conditions, accounted for all observable facts. Sometimes, when faced with specific events, these explanations became a bit clumsy. How, for example, account for the appearance of the products of electrolysis at the poles of a voltaic cell? If a water molecule is split into hydrogen and oxygen, how does each gas get to its respective pole, and why do they not simply bubble up in the middle? Or how is it possible to pass acids through alkaline electrolytes without there being any chemical reaction, as Davy had done? Dozens of speculations were forthcoming—such as that which suggested that the electrical fluids carried the hydrogen and oxygen in opposite directions through the wires and deposited them at the poles—but none which could satisfactorily cover all the phenomena.

Electrostatic induction seemed simpler, as being merely the mutual attraction or repulsion (at a distance) of the electrical fluid(s) across a non-conducting medium. The fluid(s) could be ' thrown in ' or ' out ' of bodies at will and was therefore separable from ponderable matter. Ponderable matter itself could be divided into two classes—conductors and non-conductors. Two consequences of this view should be noted. If electricity is separable from ponderable matter, it ought to be possible to create an absolute charge; conductors and non-conductors must differ fundamentally on the atomic level, and the distinction with respect to electricity implies two basically different *kinds* of matter.

The electric current was nothing more than the free passage of the electric fluid(s) through conductors completely analogous to the flow of water through pipes. How such a current could give rise to magnetic effects and induce a current in another wire, however, was difficult to explain.

Faraday began his investigations on the nature of conduction as a result of his discovery of electromagnetic induction. Having found that a magnet could induce a momentary electric current in a neighbouring wire under certain conditions, the next step was to show that the magnetism of a current-carrying helix could do likewise. This in turn raised two questions: what was the electric current, and what was the state of the secondary when a steady current was passing in the primary? Faraday's attempts to answer the first question were to occupy him the rest of his life but it is interesting to see that he considered the possibility that the ' current ' was a particular static arrangement of the particles—a polarization, and not a flow. This he reluctantly rejected since a magnetic needle placed under a wire connected to one end of a voltaic pile but not to the other shows no deviation; the moment contact with the other end is made, the needle moves indicating that something has passed. Faraday's working definition of a current, however, is worth citing for it is wider than that which immediately comes to mind.

> " By current I mean anything progressive, whether it be a fluid or Electricity or vibrations or generally progressive forces."

If the current itself is not an arrangement, must it not be preceded by some kind of change in the relations of the ultimate particles? What, for example, is the condition of the secondary after the first surge of current has passed through it when the primary circuit is made? It has returned, seemingly, to its normal state; but this is only appearance for,

> " . . . it resists the formation of an electrical current in it, whereas, if in its common condition, such a current would be produced, and when left uninfluenced it has the power of originating a current, a power which the wire does not possess under common circumstance."

Attempts were made to detect this state, which Faraday clearly felt ought to manifest itself as electrical or magnetic attractions or repulsions, but to no avail.

Nevertheless, he felt able to describe the condition of the molecules of the wire; it was, he said, a state of tension, in which the constituent particles were strained along the direction followed later by the current. This electro-tonic state, as it was named, was not restricted to the secondary in the electromagnetic arrangement. Faraday immediately seized upon it for the light it shed upon a totally different area—namely upon the problem of electrolysis and the separation of the elements under electric influence.

> " The reason given with regard to metals extends also to fluids and all other conductors, and leads to the conclusion that when electric currents are passed through them they also assume the electro-tonic state. Should that prove to be the case, its influence in voltaic decomposition, and the transference of the elements to the poles, can hardly be doubted. In the electro-tonic state the homogeneous particles of matter appear to have assumed a regular but forced electrical arrangement in the direction of the current, which if the matter be undecomposable, produces, when relieved, a return current: but in decomposable matter this forced state may be sufficient to make an elementary particle leave its companion, with which it is in a constrained condition, and associate with the neighbouring similar particle, in relation to which it is in a more natural condition, the forced electrical arrangement being itself discharged or relieved, at the same time, as effectually as if it had been freed from induction. But as the original voltaic current is continued, the electro-tonic state may be instantly renewed, producing the forced arrangement of the compound particles, to be as instantly discharged by a transference of the elementary particles of the opposite kind in opposite directions, but parallel to the current."

Only a few months later, further experiment on electromagnetic induction caused Faraday to reject the electro-tonic state, *but only for the secondary*, not for electrochemical phenomena. In that area this concept provided certain experimentally testable ideas. If Faraday were right, then ' action at a distance ' between the two poles could not also be right. This could be examined by eliminating poles. Thus if a static electricity machine were made to discharge through a piece of absorbent paper soaked in an electrolytic solution and into the air, electricity would pass without there being any centre of attraction towards which the elements of the electrolyte could be attracted. There was, nevertheless, electrolytic action, leading Faraday to conclude that

> " the effects of decomposition would seem rather to depend upon a relief of the chemical affinity in one direction and an exaltation of it in the other, rather than to direct attractions and repulsions from the poles, etc., etc."

A further testable difference between Faraday's view of the mechanism of electrolytic dissociation and decomposition and the commonly held view concerned the elements of the decomposing electrolyte. According to the ordinary view, the positive and negative electric forces literally tore the compound asunder, attracting the electro-positive and electro-negative elements to opposite poles. During their passage, the elements supposedly existed as free chemical agents, as shown by the intensification of the colour of an indicator as the particles converged on the pole. In Faraday's theory, the elements were passed along ' hand to hand ' as in an American square dance, so that at any point both the electro-positive and electro-negative elements were in combination but passing by one another in opposite directions. In order to decide between these explanations, the effect of diffusion of the indicator must be eliminated and the process slowed down as much as possible. The arrangement adopted by Faraday was extremely ingenious and characteristic of his ability to devise experiments capable of producing exactly those effects to be studied while excluding possible sources of error.

Faraday made a number of jellies capable of conducting a current and containing various electrolytes. He then arranged things as shown in fig. 2. (1) is the positive pole; (2) litmus paper just wetted (without running) with distilled water; (3) unsaline jelly; (4) litmus paper wetted as above; (5) jelly containing sodium sulphate; (6) turmeric paper wetted as above; (7) unsaline jelly; (8) turmeric paper wetted as above; (9) negative pole. When contact was made, both the litmus paper at (2) and the turmeric paper at (8) turned colour, indicating the presence of acid and alkali at the poles. At (4) and (6) nothing happened and Faraday concluded:

> " Hence the transference was good, ready and effective, but the evolution seemed to be altogether at the poles. No signs of the free acid or alkali in its transit appeared."

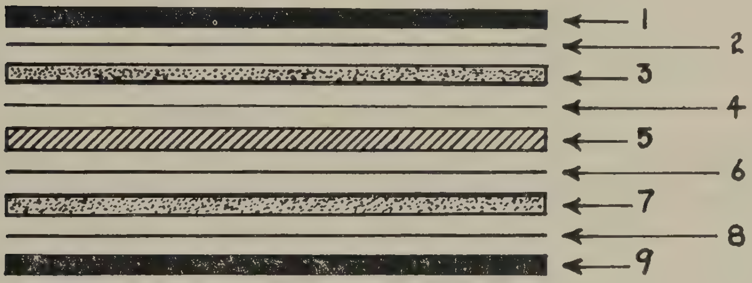


Fig. 2. Faraday's experiment on ionic migration.

Having proved, at least to his own satisfaction, that the state of tension existed during electrolysis, he now turned his attention to the whole *process* of electrolytic decomposition. The next problem was to determine what caused the state of tension, whose release and immediate re-creation constituted, in time, the electric current. We must not forget that Faraday was basically a chemist and that he saw these problems as chemical ones. As he wrote time and time again, " the forces termed chemical affinity and electricity are one and the same ".

By 1834, the theory of immaterial atoms, capable of deformation and strain by the mutual action of their surrounding forces enabled him to give a complete description of electrolysis, based on experimental facts.

> "Assuming it sufficiently proved, by the preceding experiments and considerations, that the electro-motive action depends, when zinc, platina, and dilute sulphuric acid are used, upon the mutual affinity of the metal zinc and the oxygen of the water . . . . it would appear that the metal, when alone, has not power enough under the circumstances, to take the oxygen and expel the hydrogen from the water, for, in fact no such action takes place. But it would also appear that it has power so far to act, by its attraction for the oxygen of the particles in contact with it, as to place the similar forces already active between these and the other particles of oxygen and the particles of hydrogen in the water, in a peculiar state of tension or polarity, and probably also at the same time to throw those of its own particles

> which are in contact with the water into a similar but opposed state. Whilst this state is retained, no further change occurs; but when it is relieved, by completion of the circuit, in which case the forces determined in opposite directions, with respect to the zinc and the electrolyte, are found exactly competent to neutralize each other, then a series of decompositions and recompositions takes place amongst the particles of oxygen and hydrogen constituting the water, between the place of contact with the platina and the place where the zinc is active; these intervening particles being evidently in close dependence upon and relation to each other. The zinc forms a direct compound with those particles of oxygen which were, previously, in divided relation to both it and the hydrogen: the oxide is removed by the acid, and a fresh surface of zinc is presented to the water, to renew and repeat the action."

In this explanation, note that there is one weak place in the argument; this is the existence of the electro-tonic state. It can only be inferred from the spark that will pass when the wires connecting the opposite poles of a voltaic battery are brought close together. Faraday argued that as the spark occurs before metallic contact is made and therefore before the electrolytic action could begin, there must be a pre-existent state of tension. But he would like to detect it, as it existed, not infer its presence from its effects. In May, 1833, he devised an experiment for this purpose. In view of the discovery of the rotation of polarized light in a magnetic field in 1845, the following test for internal strain, as in the electro-tonic state, deserves special attention.

> " Thinking it possible that the discharge of the electric forces by the successive decompositions and recompositions of the particles of the electrolyte might neutralize and therefore destroy any effect which the first state of tension could by possibility produce, I took a substance which, being an excellent electrolyte when fluid, was a perfect insulator when solid, namely, borate of lead, in the form of a glass plate, and connecting the sides and the edges of this mass with the metallic plates, sometimes in contact with the poles of a voltaic battery, and sometimes even with the electric machine, for the advantage of the much higher intensity then obtained, I passed a polarized ray across it in various directions, as before, but could not obtain the slightest appearance of action upon the light. Hence I conclude, that notwithstanding the new and extraordinary state which must be assumed by an electrolyte, either during decomposition (when a most enormous quantity of electricity must be traversing it), or in the state of tension which is assumed as preceding decomposition, and which might be supposed to be retained in the solid form of the electrolyte, still it has no power of affecting a polarized ray of light; for no kind of structure or tension can in this way be rendered evident."

The 1845 discovery lies beyond the scope of this article, but we cannot pass on without noting what a remarkable conclusion Faraday, the so-called strict empiricist, reached. The test showed, said Faraday, not that the electrotonic state did not exist, but that the state of tension (for which there was no direct evidence) had no power of affecting a polarized ray of light! Even so was Faraday caught in his own hypothesis despite his extraordinary sensitivity to the pitfalls of theory.

If the electro-tonic state exists in a substance which is an electrolyte when fluid and an insulator when solid, then what is it that constitutes the difference between the two states? Why, in short, should the substance conduct when fluid and not conduct when solid? There is one obvious difference between the two; in the fluid state the particles are mobile while in the solid state they are not. Conduction, therefore, in this view, is the breakdown of the electrotonic state under the tension created by the impressed electric force. But is this not precisely what happens when the tension across a so-called nonconductor becomes so great that discharge takes place? In short, are not conduction and non-conduction merely a matter of degree and not a fundamental distinction between different substances? The evolution of these ideas is again best told in Faraday's words:

> "As, therefore, in the electrolytic action, induction appeared to be the first step, and decomposition the second (the power of separating these steps from each other by giving the solid or fluid condition to the electrolyte being in our hands); as the induction was the same in its nature as that through air, glass, wax, etc., produced by any of the ordinary means; and as the whole effect in the electrolyte appeared to be an action of the particles thrown into a peculiar or polarized state, I was led to suspect that common induction itself was in all cases an action of contiguous particles, and that electrical action at a distance (i.e. ordinary inductive action) never occurred except through the influence of the intervening matter."

Once again, there was a clear-cut and testable difference between Faraday's and the commonly-held theory: induction, as action at a distance, should take place along straight lines; induction, as the result of the action of contiguous particles under tension, should take place in curves. The model was drawn once more from electrochemistry, since the deposition of elements uniformly on the electrodes clearly proved action in curves, not in straight lines. Mapping the ' lines of force ' in an enormous number of experiments Faraday was able to show that induction *did* take place in curves. This, to him, was conclusive.

It also led him to one final result, the discovery of specific inductive capacity for,

> " if induction be an action of contiguous particles, and also the first step in the process of electrolyzation . . . . there seemed reason to expect some particular relation of it to the different kinds of matter through which it would be exerted, or something equivalent to a specific electric induction for different bodies, which, if it existed, would unequivocally prove the dependence of induction on the particles."

The seeming differences in the electrical properties of bodies could now be explained in terms of the capacity for deformation of the contiguous particles out of which these bodies were composed. The metals could sustain only a very low level of tension, so that a very small deforming force caused the rise and almost immediate collapse of the electro-tonic state. The metals, in short, were good conductors. In electrolytes, the strain led to the decomposition of the compound, relief of the strain coinciding with the liberation of a particle of the elements at both poles. In insulators, the particles were able to withstand much greater tensions but here, too, discharge could be effected by placing such strain on the linked particles that the chain broke and the polarized state collapsed.

We have now reached the end of our examination of this phase of Faraday's work. By 1837, he had brought all electrical phenomena under the universal law of induction. Is it not paradoxical that Faraday, cited constantly as the foe of theorizing in general and of atomism in particular, should have built so great an edifice upon the theory of immaterial atoms?

REFERENCES

FARADAY, M., 1816–19, *A Course of Lectures on Chemistry delivered before the City Philosophical Society* (Manuscript: Institution of Electrical Engineers Library); 1844, 1855, *Experimental Researches in Electricity*, 3 vols. (London: Taylor & Francis); 1859, *Experimental Researches in Chemistry and Physics* (London: Taylor & Francis).

LAVOISIER, A., 1864–69, *Oeuvres Complètes*, 6 vols. (Paris: Imprimerie Impériale).

MARTIN, T., ed., 1932–36, *Faraday's Diary*, 8 vols. (London: G. Bell & Sons).

Science in Space

by I. M. STEPHENSON
University College, London.

A Review of the Thirteenth Conference held by the International Scientific Radio Union at University College, London from September 5 to 15, 1960.

1. HISTORICAL

Radio wave propagation had a surprising start in that it was predicted by some extremely gifted mathematicians in the middle of the 19th century.

The theoretical work on radio waves was initially correct and showed that these waves travel in a straight line through a homogeneous medium, that is a medium whose electrical properties are virtually constant and independent of the direction in which the waves are travelling. Experiments with very high frequency electromagnetic waves in the form of light had already proved this theoretical work.

At that time it was realised that the earth's atmosphere did not extend to an infinite height and it was supposed that at a great distance from the earth there was only a virtual vacuum. This would not prevent radio signals radiated from the earth from travelling out into space. As the earth is round it was assumed that any radio signal transmitted from the earth could only be received at a distant station if there was direct visibility (an 'optical path') between the two stations. This would place a severe limit on the distance between the transmitter and the receiver unless both were placed on extremely high towers. Marconi's great faith in the use of radio prompted him to try an experiment in which he placed the receiver well beyond the limit of sight, on the other side of the Atlantic; although the experts predicted that radio transmission would be impossible, he succeeded in sending signals across the Atlantic. For a time it looked as if the theoretical work of Maxwell and others was incorrect but the reason for this apparent anomaly was soon discovered. The atmosphere near the earth was found to be far more complicated than had been thought. Instead of the simple atmospheric layer followed by a vacuum there are various layers that reflect radio waves and cause them to travel round the world in a series of hops in which they are alternately reflected from the atmosphere and the earth. When attempts were made to establish regular communication networks it was found that the reflections from the various layers were very unreliable and that the conditions for satisfactory radio transmission over long distances could not be predicted. And URSI was founded in 1919 to discuss and study these communication problems.

The sun's rays cause various gases that form the upper atmosphere to become ionized and it is these ionized layers that mainly reflect the radio waves. The intensity of the radiation from the sun varies from day to day, from month to month, and from year to year quite apart from its variations in different parts of the world. The formation of this *ionosphere* is even further complicated by the fact that the air is composed of not one gas but a mixture of gases and the effect on each gas of the sun's radiation is different. An international organization such as URSI (which meets every three years) was essential to co-ordinate the work of the many research stations all over the world. It was soon realised that the propagation of radio waves gave clues to the very complicated composition of the space surrounding the earth. Although for many years astronomers had been studying the universe by examining all the visible phenomena, the advent of radio waves was enabling man to study many of the *invisible* phenomena that occur continuously all around our own planet. It was found, for instance, that the magnetic field of the earth, relatively weak though it is, has a strong influence on the composition of the ionosphere.

The wavelength λ frequency f and velocity c of electromagnetic waves are related by the expression $\lambda = c/f$. In a vacuum the velocity is about 3×10^8 metre/sec and in air it is substantially the same. In order to propagate an electromagnetic wave efficiently the aerial has to be comparable in size with the wavelength used. Early work on long distance communication was carried out at frequencies ranging from about 20 kilocycles/sec to a few megacycles/sec, that is wavelengths varying from a few tens of thousands to a few hundred metres. The early valves would not operate satisfactorily at higher frequencies.

As both valves and circuits improved so the radio transmissions were carried out at higher and higher frequencies. When short wave radio came into use it was found that the reflecting layers surrounding the earth behaved very differently at these higher frequencies, and if the frequency was high enough the radio waves were not reflected at all. Thus after some twenty years the original predictions of the theorists were to come true and at these higher frequencies communication was only possible over an optical path. Today we are quite familiar with the very high frequency microwave links that are used to transmit television programmes all over Europe, and these links have aerials placed on high towers. Even so, their range is severely limited by the curvature of the earth. Although this change in the reflecting properties of the ionosphere is a disadvantage from some points of view it does enable us to send out radio signals beyond our own atmosphere.

During the second world war, the development of metre wavelengths, and then of centimetre wavelengths, enabled a whole array of new uses to be found for radio waves. To understand why they depend on the improvement in high frequency techniques we must again return to the relationship between frequency (or wavelength) and aerial size. An aerial which is about the same size as the wavelength will radiate efficiently but will tend to radiate the waves in all directions. In order to make the aerial directional it has to be quite large compared with the wavelength. Thus although we may make aerials efficient at wavelengths of a hundred metres or so it is not practicable to make very directional aerials as they would have to be many thousands of metres long. Furthermore a directional aerial has to have a considerable area and can no longer be a long single wire. At very short wavelengths it is possible to make aerials that are large compared with the wavelength and consequently highly directional. This enables radio echo, or radar, techniques to be used in which a thin beam of radio waves is used to explore the sky. Any obstacle in the path of this beam will reflect a small part of it which will then be returned to the aerial where it can be detected by a sensitive receiver.

The shortest radio waves that can be generated by modern valves have a wavelength which is only a millimetre or so. The longest visible light waves have a wavelength of about 7×10^{-4} millimetres so that the gap in the frequency spectrum between radio waves and visible light is gradually being closed. For this reason many of the aerials used at very short wavelengths resemble the mirrors used for optical work as, of course, they are both used for focussing electromagnetic waves, the only difference between the two being the frequency of the waves.

At the end of the war the development of microwave valves and circuits, aerials, sensitive receivers and radar techniques set the scene for another major development in science, the exploration of space. One of the startling discoveries soon made was that there are a number of both visible and invisible bodies in the universe that actually radiate radio waves, as distinct from visible electromagnetic waves. In fact, the advances in radio enabled astronomers to discover new stars which had remained undetected by the optical telescopes.

The really startling event that expanded the whole scope of space radio was the improvement in rocket research which enabled satellites containing instruments to be placed in orbit so as to make measurements of conditions in space and transmit the information by radio back to earth. Up to this time we had only been able to ' look out " from the earth with our radio telescopes but the new satellites enabled scientists to both look in at the earth from outer space and to look out into space without the limitations imposed by the earth's atmosphere.

2. The 1960 Conference

The conference was arranged to cover seven different subjects, each under a separate Commission. It was opened by Dr. L. V. Berkner, the President of Associated Universities, Inc., who has been President of URSI since 1957.

The work of *Commission I* on standardizing measurements of radio frequency power, time and other essential quantities may seem rather mundane. However when attempts are being made to fire rockets at speeds of some 18 000 miles per hour into our very complicated system of planets, all of which are moving relative to one another and all of which exert gravitational pulls on the rocket, precise timing is essential if the rocket is ever to reach its destination. It is also necessary for the rocket itself to carry an accurate time standard to control its own flight.

The quartz crystal clock, which uses the electrically stimulated mechanical oscillations of a thin quartz plate, was favoured for many years but recently the ' atomic ' clock, which uses the resonance of caesium atoms, and has proved more accurate and stable than the quartz clock, has replaced it for many uses. These new standards of time were discussed, as were also the international time standards which are broadcast so that observers all the world over may have a common time standard in order that they may compare their measurements accurately even though they were taken at many different points on the earth's surface.

Another subject for discussion was the measurement of power at radio frequencies, by converting it to heat and measuring the rate of heat production or by measuring the radiation pressure exerted by the electromagnetic wave on a metal vane inside a waveguide.

Commission II studied the propagation of radio waves through the troposphere which is relatively dense and also contains water vapour which considerably modifies its propagation characteristics. Studies by meteorologists have helped to explain some of the phenomena observed in radio wave propagation and, more recently, radio observations of the troposphere by radar devices have provided meteorologists with advance warning of storms and other weather changes.

The demand for more and more communication channels and the shortage of available frequencies has led to the investigation of a new means of transmission known as the tropospheric scatter propagation. In this, a powerful radio transmitter operating at a few hundreds of megacycles a second is used and the aerial is directed not towards the distant receiving station, but up at the sky (fig 2).

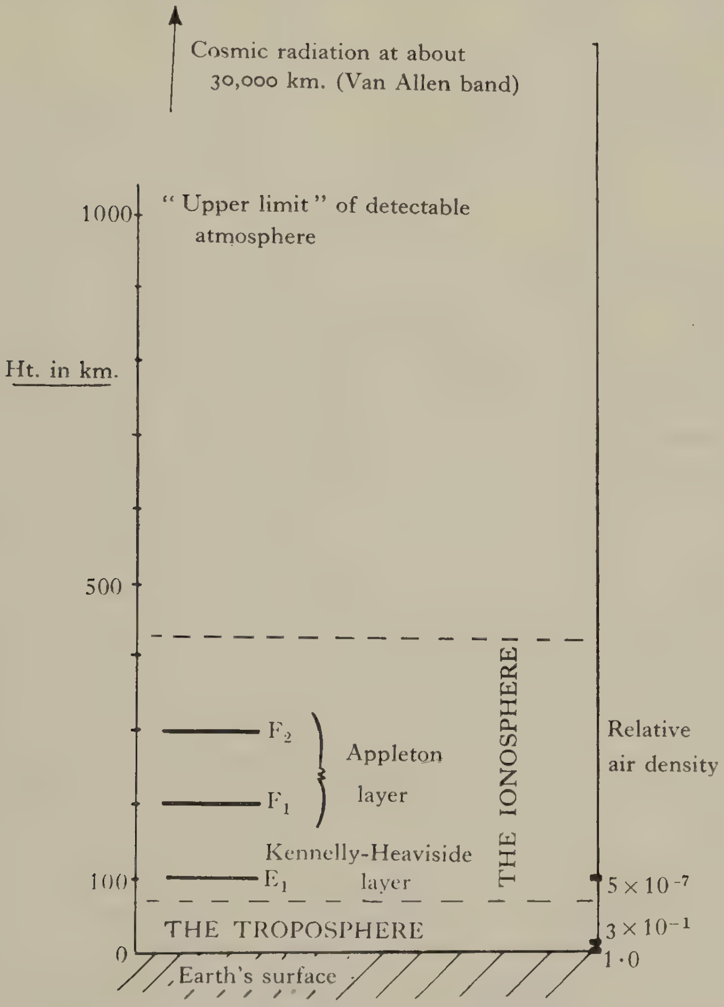


Fig. 1. The atmosphere surrounding the earth.

At such frequencies transmission would normally only be possible over an optical path as the waves would not be reflected from the ionosphere but it is found that a small proportion of the transmitted signal is scattered by the troposphere and weak, but useful, signals can be received at a distant point well beyond the horizon. The system is a wasteful one in that only a small proportion of the transmitted power is scattered in the right direction but nevertheless it is considered worthwhile operating such systems to obtain extra communication channels.

During the conference the proposals for using space satellite relays for long distance communication were compared with those for tropospheric scatter propagation, and it would appear that the former system may be the more useful as it does not require such large transmitted powers. Furthermore, the scattering process in the troposphere varies very considerably with the weather conditions. Initial experiments with the American space relay system using the 'Echo' balloon indicate that this system may be more reliable.

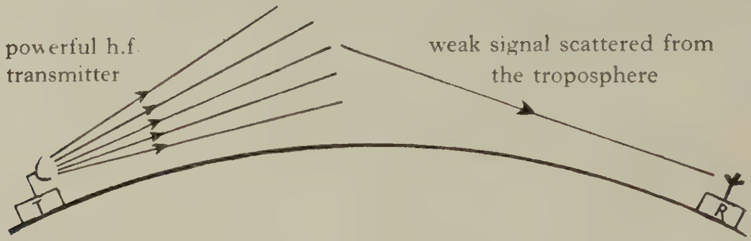

powerful h.f. transmitter
weak signal scattered from the troposphere
T
R

Fig. 2. Principle of tropospheric scatter propagation.

Commission III studied the propagation of radio waves through the ionosphere. The result of solar ionization is to produce a layer whose refractive index differs very much from that of a vacuum so that electromagnetic waves at fairly low frequencies are refracted and, in the case of most of the lower frequency waves, returned again to earth. Radio waves with frequencies of 30 Mc/s and above are not seriously affected by the ionosphere and will travel straight through it. The reason why the effect of the ionosphere varies with the frequency is apparent if we consider in a rather crude way what happens when the electromagnetic wave enters the ionised region. In this region there are charged particles which will be affected by the electric field of the radio wave and this alternating electric field will cause forces to be exerted on the charged particles. This does not occur in an un-ionized gas as the particles are then electrically neutral. As the electric field is alternating very rapidly the electrons and ions will try to move very rapidly and, due to their finite mass, the extent to which they move will decrease as the frequency of the applied field increases. Owing to their relatively large mass the movement of the ions is small enough to have no noticeable effect but the movement of the electrons is large enough to alter the transmission properties of the ionosphere. At very high frequencies, however, even the electrons do not move far enough to cause any noticeable effect and so the ionosphere once again has properties of transmission similar to that of a normal un-ionized region.

The International Geophysical Year 1958-59 has enabled an enormous amount of data concerning the ionosphere to be collected and systematically analysed. One of the factors influencing the ionosphere are sunspots, and the IGY was arranged to occur during a sunspot maximum cycle. Results from the measurements made during this period show that it would be extremely interesting to conduct similar measurements during a sunspot cycle minimum and the conference have made arrangements to co-ordinate these measurements.

Measurements on the ionosphere have been greatly aided by rockets, and one of the interesting results of such measurements is that the earth's atmosphere, at least in a highly rarefied condition, continues up to heights of at least 1000 km. whereas the upper limit was previously taken to be about 400 km. Other measurements in rockets have detected the cosmic radiation mentioned elsewhere.

Commission IV discussed radio noise and concerned itself with both naturally produced radio noise and man-made noise from electrical machinery and apparatus. Whilst the former provides the astronomer with much useful information about processes occurring fairly near the earth's surface the man-made noise is a serious nuisance and may prevent any measurements from being made. Unfortunately it is difficult to prevent unwanted radiation from electrical machinery and apparatus but URSI tries to prevent any serious interference to scientific work due to unsuitable allocation of frequency channels. Radio astronomers are also rather worried by the proposals for space satellite relays which involve placing in orbit a large band of tiny dipoles which would strongly reflect radio signals. It is feared that these dipoles might also cause serious interference with radio telescopes and it is hoped that URSI will enable agreement to be reached between the commercial interests in space communications and radio-astronomers.

Naturally produced noise comes from many sources and its study during the IGY has been most profitable. Many measurements have been made on ' whistlers ' which are of more serious scientific consequence than their name might imply. ' Whistlers ' are heard simply by connecting a long aerial to an audio amplifier and appear as a high-pitched signal starting at about 5 kc/s. and descending slowly in frequency. These whistlers are the result of lightning flashes which produce strong bursts of radio frequency power at about 5 kc/s. The propagation of the resulting wave round the earth's atmosphere causes a received signal that appears to vary in frequency in a manner that depends on atmospheric conditions.

Commission V was concerned with radio-astronomy. Large instruments such as the steerable telescope at Jodrell Bank are used only for special purposes, and much of our present knowledge has been, and is still being, collected by much smaller instruments all over the world. The titles of such papers as ' Observations of Jupiter ' and ' Centimetre wave radiation from Venus ' give some idea of the work discussed by the commission.

Many of the telescopes used have large parabolic mirrors which are rather similar to their optical counterparts but another instrument used widely is the radio interferometer. This has conventional aerials which are fixed to the ground fairly near each other. The device is called an interferometer because the signals from the various aerials are combined together and the resultant output depends on the relative phases of the signals received by the different aerials. The phases of the signals in the aerials depend on the distance they have travelled from the source (a radio star perhaps) to each aerial and as the aerials are spaced apart these distances, and consequently the phases of the signals, are different. As the earth rotates the aerials scan the sky and the phases of the signals from the aerials vary ; if the signals tend to be out of phase with each other so the total output from all the aerials will become less, similarly if they tend to all be in phase the output will be a maximum.

Another large radio telescope now under construction has a huge parabolic mirror cut out of solid rock. Here again the rotation of the earth enables it to scan the sky.

Radio astronomy has greatly increased our knowledge of the universe and has not only provided more information about those parts already observed but has shown that the universe is considerably larger than we had previously imagined. It had been realised for some time that the temperatures of the stars gradually decrease as they age and that the frequency of the electromagnetic radiation from them likewise decreases. Thus it was assumed that there must be many stars that were once visible but which are now no longer observable by optical means as the frequency of their radiation lie now lies within radio, rather than optical frequency bands. It is interesting to note that various bodies that now emit radio signals were observed by optical means in 1572 and 1604. Since then they have ceased to emit visible light and have been lost. Thus after an interval of over 400 years they have been re-discovered and the observations of the very early optical astronomers verified.

Commission VI considers radio wave propagation under rather special conditions. One of these was the propagation of waves close to a conducting surface coated with a dielectric material. Another subject studied was propagation in inhomogeneous media and an experiment was described in which a number of balls made of very light material were kept moving in the air by an air jet and radio waves were propagated through the mass of moving balls. This large scale model simulated the effect of moving air molecules in a turbulent atmosphere.

Probably the most spectacular meeting of this Commission was the joint meeting with all other Commissions to discuss the possibility of space relays. One method of transmitting radio signals from one side of the earth to another is to direct radio waves into space so that they are reflected from an artificial satellite and return to earth at some distant point. This is a passive relay and is shown in fig. 3. A more advanced system proposed is to use an active satellite which contains an amplifier so that the signals received by the device are re-radiated with considerably increased power. The amplifier would use a travelling wave tube and the satellite would have to contain power supplies for the various valves (fig. 4).

Experiments already conducted with the 100 ft American balloon have produced results that have surpassed expectations, and commercial firms have already estimated the economic possibility of such systems. Although the cost of putting an active satellite into orbit is very high, the income would rapidly cover the cost. The satellite would contain tape recorders that would record messages on one side of the earth and retransmit them at any suitable time when ‘ commanded ’ to do so by a radio signal from the earth.

Another scheme suggested is the ‘ Needles ’ project in which thousands of tiny dipole aerials would be placed in orbit. Resonant dipoles reflect radio waves of their resonant frequency very strongly and, for a given area of metal, are equivalent to quite large reflecting surfaces. As the dipoles would gradually fall out of their correct orbit and become useless they would be made of tin. When they fell out they would enter very cold regions and the tin would turn from its normal form to powdery grey tin, so that the device would automatically destroy itself when it became useless.

Commission VII studied the specialized electronic equipment necessary for radio-astronomical and satellite apparatus.

The radio power emitted from most radio stars is microscopically small and has to be amplified by some means before it is measurable. Unfortunately most ordinary amplifying valves which amplify by using streams, or beams, of electrons themselves produce noise which may be very much stronger than the signal from the radio star.

This does not *prevent* observation of the signal but its extraction from the background noise is a difficult task. Two new devices, which give much less inherent noise, are parametric amplifiers and masers. In the first system a variable parameter, which may be a variable capacitance, inductance, or a rotating electron beam of variable diameter of rotation, is used in a special circuit. The basic form of the maser is shown in fig 5. Both these devices have recently been used to provide low noise amplifiers for radio telescopes. Although masers are fairly widely used at the moment they have a disadvantage in that many forms will only operate efficiently at very low temperatures and have to be cooled with liquid helium. As the amplifier has to be attached to the aerial of the radio telescope and this aerial often is moveable the provision of the liquid helium cooling jacket sets various practical problems and it is hoped that parametric amplifiers, which will operate satisfactorily at room temperatures, may prove more useful in practice.

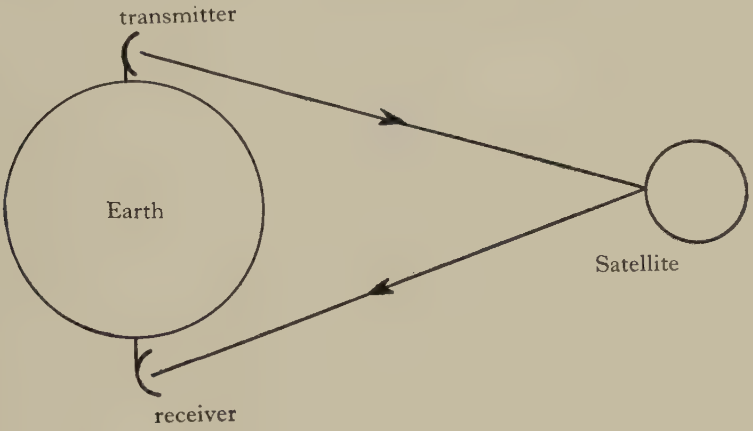


Fig. 3. Passive satellite relay.

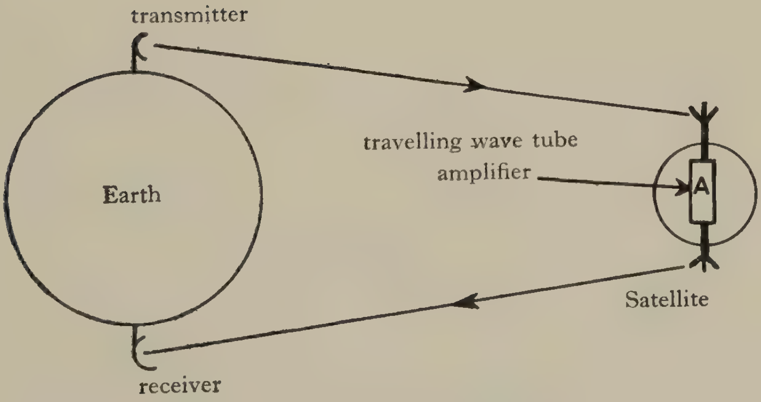


Fig. 4. Principle of active satellite relay.

One of the great problems in a space vehicle or artificial satellite is to provide the power. Batteries are heavy and short-lived, but the radiation from the sun can be harnessed to provide a permanent power source. At present the most useful generators are semi-conductor photo cells or thermoelectric devices. The efficiency of these at the moment is only about 10–15% and although this efficiency is low it is quite satisfactory for providing power in satellites where relatively large solar powers are available. Another possible means of converting heat directly into electricity is to utilise the small emission that comes from a hot cathode.

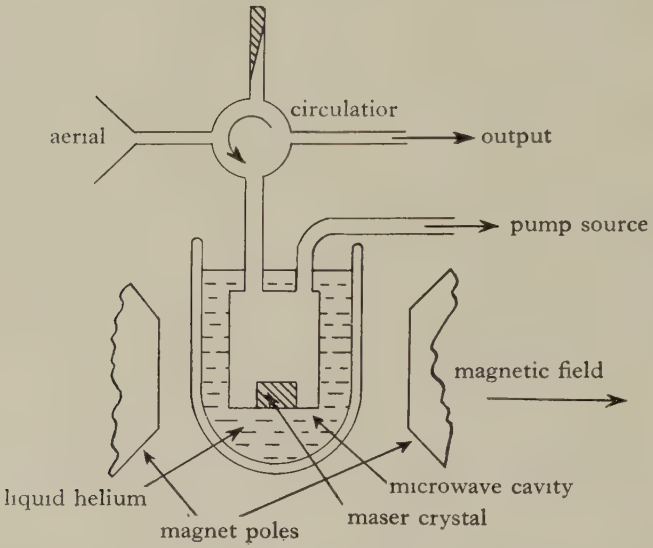


Fig. 5. The microwave solid state maser. The microwave cavity is the very high frequency equivalent of a tuned resonant circuit. It contains a maser crystal, and power from the pump source (a microwave oscillator) is fed into this cavity. Some of this power from the pump source is transferred to the crystal and temporarily stored there. A signal entering the aerial passes to the circulator, which is a uni-directional device, and enters the cavity where some of the energy stored in the crystal is transferred to it. The amplified signal passes back out of the cavity, and the circulator allows it only to pass along the output waveguide which will be connected to the next stage of amplification. The crystal in this type of maser has to be kept at a very low temperature and also requires a strong magnetic field.

I am very grateful to my colleague Dr. G. D. Sims for many helpful suggestions during the preparation of this report.

The Author:

Dr. I. M. Stephenson is a lecturer in Electrical Engineering at University College, London. He was educated at Highgate School, Imperial College, and University College. He is chiefly interested in microwave electronics, on which he has published a number of papers.

Corso di Aggiornamento

An International Refresher Course for Physics Teachers held at the 1960 Conference of the Italian Physical Society

The forty-sixth Conference of the Italian Physical Society, held in the Mostra d'Oltemare at Naples from September 29 to October 5, 1960, was accompanied by a Refresher Course for physics teachers which was organized in collaboration with the Organization for European Economic Cooperation (OEEC), and the Italian Ministry of Public Instruction. As Lord Hailsham emphasized in his speech at the Brussels Seminar in April, economic progress can only follow in the wake of scientific education, which is indeed a first priority with OEEC. The draft programme of its Council, issued in March 1959, outlined plans for the development of basic education in science by bringing syllabuses into line with the present climate of scientific thought, running courses for teachers and seminars for the study of general problems, exchanging personnel at all levels, and sending out teams of experienced teachers to the less developed countries. It also said that most of this would have to be postponed because of funds, but the right sort of shoestring seems to be available now. The Naples Course, arranged by Professor G. Polvani of Milan (President of the Italian Physical Society) and Professor A. Alligretti of Pisa (OEEC), was attended by about fifty Italian teachers and thirty from other countries.

The lecture programme included: " Principles of Classical Mechanics ", " Relativistic Mechanics ", and " Relativistic Optics and Electrodynamics ", by S. Sikjaer (Copenhagen); " Fundamentals of Hydrostatics ", by J. P. Stephenson (City of London School); " Geometrical Optics ", " Methodology of Optics ", and " Twentieth Century Physics ", by H. F. Boulind (Cambridge University); " Fundamental Experiments in Electrostatics and Electrodynamics ", by G. R. Noakes (Uppingham School); " Thermodynamics ", by G. Cortini (Naples); " Physical Optics ", and " The Methodology of Mechanics and Electromagnetism ", by A. Carrelli (Naples); " Solid State Physics ", by W. Dekeyser; and " Thinking Machines ", by E. Caianiello. A. Rostagni reported on the Paris Conference (July 1960) of the International Union of Pure and Applied Physics on Physics Education (at which " Paper 40 ", with its suggestions for new syllabuses for school physics teaching, had been presented by Norman Clarke); and B. Rossi gave a lecture-demonstration of the excellent apparatus devised by the Physical Science Study Committee which had been exhibited at this Conference. Representatives of Phywe (Göttingen), Leybold (Cologne), and Paravia (Turin) lectured on their teaching apparatus.

I was told that in Italy, quite apart from salaries, the main obstacle to recruiting young men into science teaching is the traditional low social prestige of the profession. A Course on this scale, accompanying the Conference of an august scientific society, must do more than any amount of exhortation or fair words could ever do to make people feel that science teaching is a calling held in high regard, that could even be commended to an able-bodied man.

G. R. Noakes.

Radar Astronomy

by J. V. EVANS
Massachusetts Institute of Technology
Formerly of the University of Manchester

SUMMARY

Radar Astronomy is a new and growing branch of Astronomy. Although it seems that radio echo studies must be confined to the solar system, they can play an important part in developing our understanding of the Sun and the planets. At the present time these objects are barely detectable by radar techniques and much of the work has been concerned only with the moon. However it seems likely that within a decade we shall be able to measure the distances of the nearer planets and study their surfaces by this astronomical technique. This paper outlines what has already been achieved in the study of the moon and indicates some of the possible future results.

1. INTRODUCTION

A new branch of Astronomy has gradually developed in recent years. Radar Astronomy can claim to have existed since 1926 when Breit and Tuve first studied the earth's ionosphere by radio-echo techniques, but perhaps a better starting point would be 1946. In that year the first radio echoes from the moon were obtained and systematic radio-echo studies of meteors began. Since then a great deal has been learned about both the radio reflection properties of the moon and the distribution, orbits and velocities of meteors. In addition the first successful experiments to detect echoes from the sun and the planets have been carried out. At the present time several large radar equipments are in the course of construction and in about three years time there should be three or four systems which will be capable of detecting radio echoes from the sun or the nearer planets.

This paper attempts to provide an outline of the experimental results which have been obtained so far. The meteor echo studies have been excluded deliberately, because Lovell (1954) has provided a comprehensive survey of this work. Most of the experimental work has been concerned with the moon and only the experiments which appear to have made large contributions to our understanding of the moon's reflection properties will be described. The list of references at the end, however, includes most of the work published up to the present time. A description of experiments to detect radio echoes from Venus and the sun is given. The paper concludes by reviewing the results which may be obtained in the next few years, and briefly describes the experiments which have already been planned.

2. THE ADVANTAGES OF RADAR

In both radar and radio astronomy it is difficult to achieve the high angular resolution which is usual in optical studies. This is because the resolving power of a telescope is given by the ratio of the diameter of the objective lens or the reflecting mirror to the wavelength of the electromagnetic waves. Thus at radio wavelengths it becomes exceedingly difficult to achieve a resolving power even comparable with that of the human eye. There are for instance few radio telescopes yet in existance which can examine the surface of the moon in any but the crudest detail.

In radio astronomy this disadvantage is offset by the ability to study regions of space which are not available to the visual observer. because radio waves unlike light waves are not attenuated by the existence of large numbers of dust particles within our own galaxy. For radar studies, however, the intensity of the signals is determined by the distance of the target raised to the fourth power (eqn. (8) below) and hence it seems unlikely that radar observations will ever be made of objects which are beyond the confines of the solar system. What then has the radar method to offer which other astronomical techniques have not?

The greatest advantage that radar provides is the ability to perform controlled experiments. The frequency, time and duration of the transmissions are known, and in consequence measurements can be made of range (from the echo-time delay) and velocity (from the doppler shift). For example the range of the moon can be measured to a greater accuracy by radar than optically (to a distance of about half a mile) and the velocity of a meteor established with greater precision. Radar can also be used to study objects at wavelengths where their own radiation is too weak to be detected. The existence of charged particles can be inferred either because they reflect the signals (e.g. the aurorae, the ionosphere, and the sun) or because they influence the propagation of the radio waves between the transmitter and another target. Finally radar provides a new method of studying the surface features of the planets, as radio waves are unaffected by the presence of obscuring planetary atmospheres. Thus in principle it will be possible to determine the rotation period of Venus, and also some of the properties of the surface, such as the electrical constants of the material and the overall surface roughness. The way in which this information can be obtained will be discussed next.

3. THEORY

3.1. *The Radar Equation*

In a radar system the radio waves are transmitted usually in the form of pulses, and these are scattered back by obstacles or targets lying along their paths. If the peak power in the pulses is P_t and the gain of the transmitting aerial system over an isotropic radiator in the direction of the target is G then the flux density at a target R metres distant is

$$\frac{P_t G}{4\pi R^2} \text{ watts. metres}^{-2}. \tag{1}$$

The target will intercept an amount of flux proportional to its projected area and some fraction of this power will be absorbed depending upon the electrical properties of the surface. The remainder of the power is reflected back in directions governed by the shape of the target, and the flux density at the observer may be written

$$\frac{P_t G \sigma}{(4\pi R^2)^2} \text{ watts. metres}^{-2} \tag{2}$$

where σ is the scattering cross-section of the target. This will be a function of the aspect from which the target is viewed and the wavelength of the incident radio waves. The scattering cross-section σ may be defined as the projected area of an isotropically reflecting body which is perfectly conducting and returns the same flux to the observer as the actual target. Where an aerial system having an aperture A metres^2 is used to receive the reflected radio waves the power P_r presented to the receiver input terminals is

$$P_r = \frac{P_t G A \sigma}{(4\pi R^2)^2} \tag{3}$$

If the same aerial is used for both transmitting and receiving the radio waves, and this is in the form of an aperture over which the transmitter power is uniformly distributed, G and A are related by

$$G = \frac{4\pi A}{\lambda^2} \tag{4}$$

where λ is the radio wavelength. In practice where parabolic or similar reflectors are used, it is extremely difficult to arrange for the transmitter power to be uniformly distributed over the aperture without also spilling over the edges. Some improvement can be made by employing an aerial system having many radiating elements each backed by a separate (or continuous) reflector, but this tends to increase the power lost in the feeder arrangements. Thus most aerial systems have efficiencies η lying between 50 per cent and 75 per cent, and the value of A in eqns. (3) and (4) is the effective aperture of the system, i.e.

$$A = \eta A^1 \tag{5}$$

where A^1 is the geometrical area of the aperture.
Thus from (3)

$$P_r = \frac{P_t A^2 \sigma}{4\pi \lambda^2 R^4}. \tag{6}$$

This received power must be amplified to work a recording device, and will then have to compete with the noise power already present at the receiver input terminals amplified by the same amount, and that produced in the amplifying stages themselves. In a well designed receiver nearly all the noise at the output terminals is caused by the noise received by the aerial system, together with that developed in the first amplifiying stage of the receiver. This latter component of the noise may be regarded as being present at the input terminals and the amplifying stage itself as introducing no additional noise. When the input terminals of a perfect amplifying stage are connected to a matched resistor at room temperature T_o the noise power present is $\boldsymbol{k} T_o b$ watts ($\boldsymbol{k}$ is Boltzmann's constant and b is the energy bandwidth of the receiver). A practical amplifying stage will have an input noise power under these conditions of $n \boldsymbol{k} T_o b$ watts, where n is the receiver noise factor and is defined by this equation. Thus when the receiver is connected to an aerial system at a temperature T_A the input noise power P_n becomes

$$P_n = \{(n-1)T_o + T_A\} \boldsymbol{k} b \text{ watts} \tag{7}$$

where $T_o = 290°K$, and T_A is the brightness temperature of the sky as seen by the aerial system. T_A is a function of both frequency and direction, and the maximum and minimum values are plotted in fig. 1 as a function of frequency. If the aerial system is directed towards a strong source of radio waves such as the sun, the centre of the galaxy, or the radio sources in the constellations of Cygnus and Cassiopeia, then the sky temperature T_A may well be greater than shown in fig. 1.

The ratio of the signal power P_r to the noise power P_n at the output of the receiver is

$$\frac{P_r}{P_n} = \frac{P_t A^2 \sigma}{4\pi\{(n-1)T_0 + T_A\} k b \lambda^2 R^4}. \tag{8}$$

Since the noise amplitude shows large random fluctuations the amount of information concerning the target which can be obtained from the echo is directly proportional to the signal to noise ratio P_r/P_n. At low values of P_r/P_n the very existence of the target can only be expressed on a probability basis.

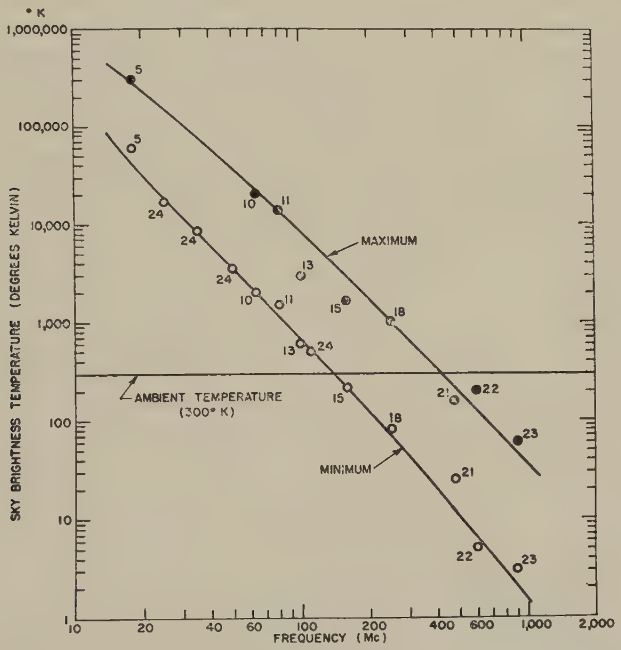


Fig. 1. The variation of the sky temperature as a function of frequency (after Kraus and Ko, 1957). The numbers adjacent to the points are reference numbers for the different observers listed by Kraus and Ko.

3.2. *The Hard Targets*

Objects such as the sun give rise to radio reflection due to the presence of electrons whose density is equal to or greater than a certain critical density. These targets must be considered separately from others such as the moon, where reflection occurs at the transition from free space to the dielectric materials of the surface. In addition to the moon the minor planets of the solar system present targets where solid surfaces exist to produce radio wave reflection. It seems probable that the major planets have solid cores, in which case they too can be classed as hard targets. The size of these cores is open to question and the amount of absorption that their dense atmospheres may introduce is also unknown. Hence it is not possible to predict the scattering cross sections σ for these planets. A very approximate estimate can be obtained by assuming that the radar cross section is equal to the geometrical cross section of the body. Then from the figures for the size and ranges of the planets given in table 1, a chart can be drawn up such as that of fig. 2 showing the relative scattering cross sections of the planets as a function of the echo delay time. The origin of this chart has been taken as a target having a cross-section of one square metre at a range of 10 000 miles. Such a target would barely be detectable with existing radar systems, but those which have already been planned for the future, should be more sensitive by some four orders of magnitude (40 dB). The moon has not been shown in fig. 2 because it would provide an echo some 70 dB greater than that from Venus. It is probable that the positions shown for all the planets in this chart are in error by one or two orders of magnitude. Certainly the moon which has been studied extensively has a scattering cross section of only about one-twentieth of its physical area, and there is evidence that Venus has a similar radio albedo. Thus if Venus, Mercury, Mars and the asteroids scatter in the same way as the moon their detectability relative to a one square metre target at 10 000 miles range, should be reduced by 13 dB. The major planets are also likely to reflect less favourably than shown in fig. 2 since it is probable that their cores will be appreciably smaller than the diameters listed in table 1, and serious absorption may occur in their atmospheres.

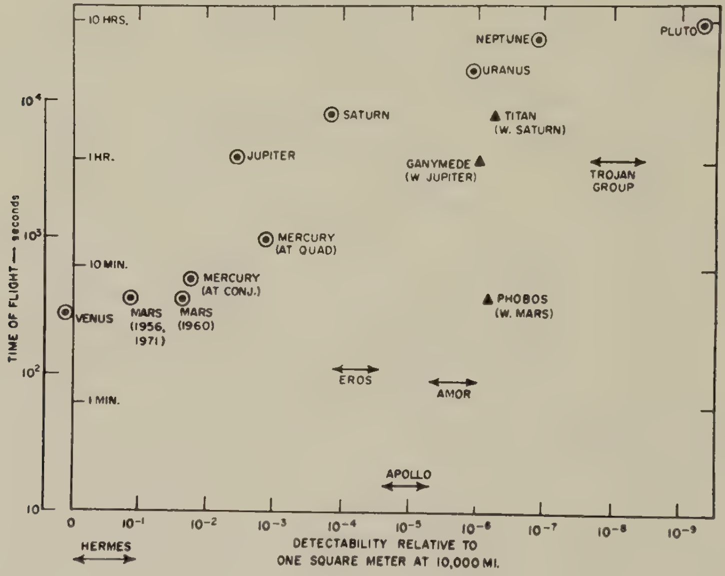


Fig. 2. The detectability of the planets plotted relative to that of a square metre target at a range of 10 000 miles. In plotting these points it has been assumed that the radar cross sections of the planets are equal to their geometrical cross sections (diagram by Green).

3.3. *The scattering cross section of a spherical target*

Where the wavelength λ of the incident radio waves is very much shorter than the radius R_0 of the target, a smooth perfectly conducting object will have a scattering cross section σ equal to the actual area of the disc (Norton and

Table 1

PLANETS AS RADAR TARGETS

Body	Closest distance (mile)	Diameter (mile)	Rotation period	Synodic period (days)	Angular size (sec.)	Number of satellites	Optically obser. temp. on the sunlit side (°K)	Critical freq. relative to earth for standard ionosphere
Mercury	$4{\cdot}8 \times 10^{7}$ (conjunction)	$3{\cdot}2 \times 10^{3}$	?	116	—	0	613	2·5
	$9{\cdot}3 \times 10^{7}$ (quadrature)	,,			9			
Venus	$2{\cdot}6 \times 10^{7}$	$7{\cdot}6 \times 10^{3}$	?	584	67	0	370	1·4
Mars	$3{\cdot}5 \times 10^{7}$	$4{\cdot}2 \times 10^{3}$	$24^{h}\ 37^{m}$	780	25	2	260	0·6
Jupiter	$3{\cdot}7 \times 10^{8}$	$8{\cdot}5 \times 10^{4}$	$9^{h}\ 50^{m}$ (equator) $9^{h}\ 55^{m}$ (temp. lats.)	399	50	11	143	0·3
Saturn	$7{\cdot}4 \times 10^{8}$	$7{\cdot}1 \times 10^{4}$	$10^{h}\ 14^{m}$ (equator) $10^{h}\ 38^{m}$ (lat. 35°)	378	20 (not including rings)	9	120	0·2
Uranus	$1{\cdot}7 \times 10^{9}$	$3{\cdot}1 \times 10^{4}$	$10^{h}\ 49^{m}$	370	3·8	5	?	0·1
Neptune	$2{\cdot}7 \times 10^{9}$	$2{\cdot}8 \times 10^{4}$	$15^{h}\ 40^{m}$ (?)	367	2·0	2	108	·05
Pluto	$3{\cdot}5 \times 10^{9}$	$3{\cdot}7 \times 10^{3}$	16^{h} (?)	367	0·23	?	?	·025

Omberg, 1947), i.e.

$$\sigma = \pi R_0^2. \tag{9}$$

If, however, the sphere is made of dielectric materials having a reflection coefficient $\rho = |Q|^2$ then

$$\sigma = \rho \pi R_0^2 \tag{10}$$

where

$$\underset{\sim}{Q} = \frac{1 - \sqrt{\left\{\frac{\mu_0}{\mu}\left(\frac{\epsilon}{\epsilon_0} + i\frac{s}{\omega\epsilon_0}\right)\right\}}}{1 + \sqrt{\left\{\frac{\mu_0}{\mu}\left(\frac{\epsilon}{\epsilon_0} + i\frac{s}{\omega\epsilon_0}\right)\right\}}} \tag{11}$$

in which μ = permeability, ϵ = permittivity, s = conductivity (the subscript o denotes free space values) and ω is the angular frequency of the radio waves. For a perfect dielectric where the conductivity $s \rightarrow 0$ and $\mu \rightarrow \mu_0$, ρ becomes

$$\rho = |Q|^2 = \left(\frac{\sqrt{\epsilon} - 1}{\sqrt{\epsilon} + 1}\right)^2. \tag{12}$$

Senior and Siegel (1960) have shown that for a smooth target the electrical constants of the surface may be determined by studying the wavelength dependence of σ. Where the target is not smooth, but has irregularities which are greater in size than $\lambda/8$, the scattering cross section becomes

$$\sigma = g \rho \pi R_0^2 \tag{13}$$

in which g is a directivity term which denotes the gain of the target over an isotropic radiator of the same size. Because a radar transmitter provides a single coherent source of radiation the gain g will be a function both of the wavelength λ and the aspect from which the target is viewed. This is because the signals returned from all the elements of the surface must be added together with regard to phase. The relative importance of the elements on the surface will depend upon the scattering mechanism—the law relating the power reflected per unit solid angle to the angular distance between the normal to the surface and the ray path. Such a law is Lambert's law which states that if the incident and reflected ray paths are the same (this is generally true for radar observation of distant objects) then the reflected power P_ϕ is related to the incident power P_i by

$$P_\phi = P_i \cos^2 \phi \tag{14}$$

where ϕ is the angle between the normal to the elemental area of surface and the ray path. A spherical target which obeyed such a law would appear limb dark. For Lommel-Seeliger scattering

$$P_\phi = P_i \cos \phi \tag{15}$$

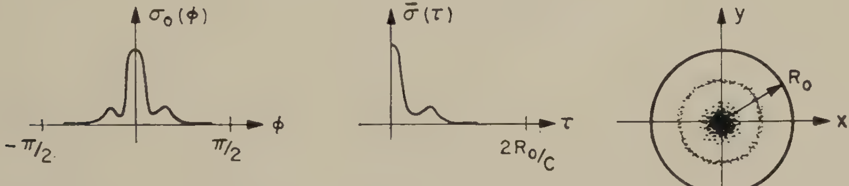
and a spherical target would then appear uniformly bright across its disc. Values of g have been derived for both laws; thus for Lambert scattering $g = 8/3$ (Grieg, Metzger and Waer, 1948) and for Lommel-Seeliger scattering $g = 5{\cdot}7$ (Kerr and Shain, 1951). Pettit and Nicholson (1930) observed that at infra-red wavelengths the moon appeared to scatter according to the law

$$P_\phi = P_i \cos^{5/3} \phi \tag{16}$$

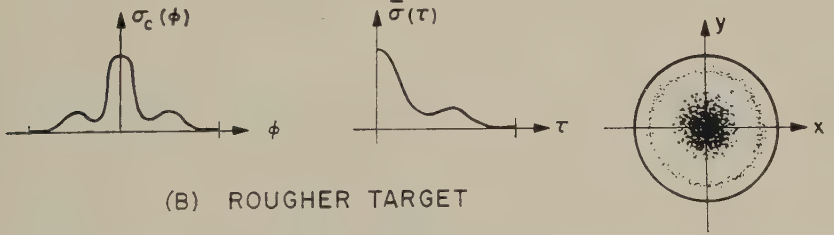
and Winter (1956) calculated the value of g for this law, and obtained $g = 2{\cdot}5$.

For a rough target a measurement of the echo intensity will provide only a determination of the product ρg. Some method must be sought to determine

ρ or g separately in order that both can be obtained. These quantities are important as they define the average electrical properties, and roughness of the surface. Because the targets are spherical the average inclination of the surface elements to the ray paths varies, and this causes a change in the surface brightness depending upon the scattering law. This is illustrated in fig. 3 for two arbitrary scattering laws. The variations of the brightness across the disc may be difficult to determine directly because of the poor angular resolution of most radar systems, but this information can usually be obtained by observing the variation of echo intensity with range, or with doppler broadening for targets which are spinning with known velocity. Where the rotational velocity is not known it may be determined by measuring simultaneously the echo intensity as a function of range and of doppler broadening.



A) RELATIVELY SMOOTH TARGET



(B) ROUGHER TARGET

Fig. 3 (a, b). The brightness distribution as a function of range $\sigma(\tau)$, shown for two hypothetical angular power scattering laws $\sigma_0(\phi)$. Also shown is the appearance of the target—the shaded areas represent bright regions (diagram by Green).

4.3. *The characteristics of the echo*

If the target is not rotating relative to the observer, then the distribution of the scattering centres over the surface can only be inferred from the distribution of echo power P_r with range R (or time). This is illustrated in fig. 4*a* for a hypothetical distribution of scattering centres. Where the target is spinning, the motion of the limbs towards and away from the observer gives rise to doppler broadening and the scattering cross section term σ becomes a function having dimensions of both time τ and frequency f. This is shown in fig. 4*a* for two values of the rotation speed W. At any instant the surface of the scattering function $\sigma(f\tau)$ will probably be more irregular than illustrated in fig. 4*a*, since the distribution of the scattering centres will tend to cause local peaks and nulls due to reinforcement and destructive interference of the signals. However by taking a time average a surface will be obtained which will represent the true mean distribution of the scattering centres and hence the scattering mechanism.

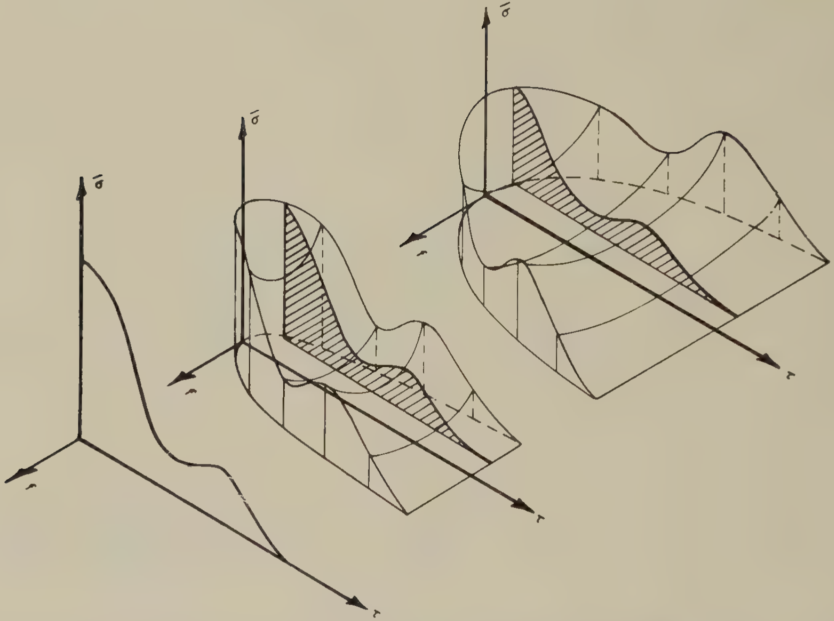


Fig. 4 a. The brightness distribution plotted as a function of range $\sigma(\tau)$ for a stationary target (lower) and above as a function of range and doppler broadening $\sigma(f\tau)$ for two targets rotating with different rotation speeds.

The accuracy with which the scattering function $\sigma(f\tau)$ can be examined depends upon the analogous function for the transmitted pulse $P_t(f\tau)$ called the abiguity function. A single transmitted pulse will be found to occupy a band of frequencies determined by the Fourier cosine transform of the pulse envelope. This means that for square pulses of length T seconds the frequency distribution of the energy is given approximately by $1/T$ cycles. This is shown in fig. 4*b* for two pulses. The echo function $P_r(f\tau)$ observed in a radar system is the convolution of the ambiguity function $P_t(f\tau)$ with the scattering function $\sigma(f\tau)$. In practice the distribution of the energy at the output of the receiver is likely to be considerably different from that at the input. This is caused partly by the finite bandwidth of the circuits and partly as a result of the detection process If the bandwidth b is made large enough the first effect can be made negligible.

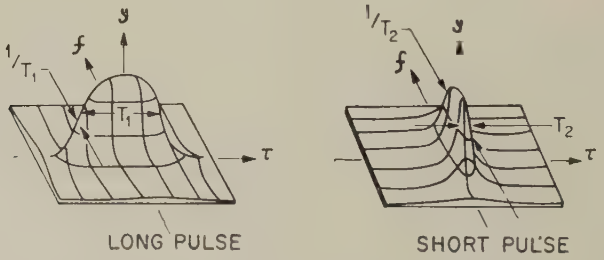


Fig. 4 b. The ambiguity function $P_t(f\tau)$ is shown for both long and short transmitted pulses (diagram by Green).

However, since the noise power in the receiver is proportional to the bandwidth this will reduce the signal to noise ratio P_r/P_n. A theorem discovered independently by Wiener, Hanson, North and Van Vleck states that the optimum signal to noise ratio is achieved when the shape of the receiver response curve is adjusted to be the same as the shape of the power spectrum of the echo. For small targets where the volume under the surface defined by $\sigma(f\tau)$ is considerably less than that under the ambiguity function $P_t(f\tau)$ the optimum bandwidth b is related directly to the length of the transmitter pulse and the criterion

$$b=\frac{1}{T} \tag{17}$$

is widely used.

For planetary echoes the greatest signal to noise ratio will be obtained when the pulse length T exceeds the depth of the target and the receiver bandwidth b is adjusted to match the spectrum of this pulse together with the broadening produced by the doppler effect. The amount of receiver bandwidth determine the ‘ rise time ’ of the signals at the output of the receiver. Thus a square pulse after amplification by the receiver will be found to have leading and trailing edges which rise and fall at a rate proportional to $1/b$ seconds.

The action of the detector is somewhat more complicated as it depends upon the ratio of the signal power to the noise power. The video signals are produced by mixing the wanted signal components with themselves, and with the noise, and mixing the noise components with themselves (see Smith 1951). Hence the voltage signal to noise ratio following the detector depends upon both the predetector power signal to noise ratio P_r/P_n and upon the law of the detector. When the bandwidth b has been adjusted to match the pulse length T, a weak echo becomes indistinguishable from the noise spikes, and as a result the observed range of such an echo will fluctuate over $\pm T$ seconds. The effects of noise can be reduced by averaging over many echoes, and in principle it is possible to apply some correction for the finite bandwidth of the receiver, but there is no correction possible to permit the true scattering function $\sigma(f\tau)$ to be derived from the observed echo function $P_r(f\tau)$ which is the convolution of $\sigma(f\tau)$ with the ambiguity function $P_t(f\tau)$. Thus to examine the scattering function $\sigma(f\tau)$ it is preferable to use pulses for which the ambiguity function $P_t(f\tau)$ contains a much smaller volume than is under the surface $\sigma(f\tau)$. In the case of Mars and probably several other planets this is quite possible, but for the moon the rate of rotation relative to the earth is so slow that the volume under the ambiguity function $Pt(f\tau)$ is always greater than that under the scattering function $\sigma(f\tau)$ (for observations at wavelengths greater than 10 cms). In this case only the distribution of the scattering centres with range $\sigma(\tau)$ can be determined (using short pulses) or with doppler broadening $\sigma(f)$ (using c.w.), unless a system of coherent pulse transmissions is employed. When this is done a high range resolution can be obtained together with a narrow frequency response. The actual power spectrum of such a chain of pulses is similar to the angular distribution of light produced by a spectrum grating. The overall envelope of the ‘ orders ’ is determined by the length of a single pulse. The width of each ‘ order ’ is determined by the total duration of the pulse train, and the ratio between the separation of the orders and their width is given by the number of pulses in the train. To reduce the size of the ‘ orders ’ on either side of the principal one the individual pulses must be as long as possible, but this is usually limited by the degree of range resolution required. An alternative approach would be to reduce the intensity of the pulses at the beginning and end of the train.

Either by this technique, or by direct examination of the scattering function $\sigma(f_T)$ using pulses of the appropriate length the average distribution of the scattering centres and consequently the scattering law can be determined. Once the scattering law for the surface is known the directivity term g can be calculated, and the reflection coefficient ρ determined from measurements of the signal to noise ratio (which provide a value for the product ρg). It must be noted that the directivity term g only has any meaning when the whole surface is completely illuminated and therefore effective in scattering at the same time. This calls for transmitter pulses which are longer than the depth of the target. Where pulses shorter than this are used the observed echoes will be weaker by an amount which will depend upon both the pulse length and the particular distribution of scattering centres for that target.

4. THE LIMITATIONS OF A RADAR SYSTEM

In the next few sections the experiments which have been carried out to study the moon and planets will be discussed. In reading these sections it is well to bear in mind the practical limitations under which the experiments have been conducted. First there are the fundamental limitations. One of these—the presence of noise in the receiver both internally generated and collected by the aerial system has been mentioned. Sites are generally chosen at which the level of man-made interference is small or negligible. The observer can do much to reduce the unwanted noise power generated in the receiver, by using special devices which have recently been developed such as the maser, the Adler tube and the reactor diode parametric amplifier. Although these devices are generally more complicated than conventional valve amplifiers they can make an enormous improvement to the noise factor n. It is now possible to make receivers which develop less noise ($(n-1)T_0 kb$) than that picked up by the aerial ($T_A kb$) for frequencies of up to about 400 Mc/s. In the very near future this upper frequency limit should be raised considerably by further developments. Beyond 2000 Mc/s the effective sky temperature (fig. 1) starts to rise again owing to atmospheric absorption. This effect together with the losses in the aerial feeder system places a practicable lower limit of about 100°K on the ambient input temperature of a receiver working in the region 400–3000 Mc/s.

The choice of radio wavelength λ is not unrestricted. Apart from international agreements concerning frequency allocations, there is a lower limit to the possible wavelength determined by atmospheric absorption, and an upper limit determined by the density of the ionosphere. As a rough guide useful observations can only be made at wavelengths lying between about 3 cms and 15 metres.

In the radar eqn. (8)

$$\frac{P_r}{P_n}=\frac{P_t A^2 \sigma}{4\pi\{(n-1)T_0+T_A\}kb\lambda^2 R^4}$$

the terms P_t, A, n, b and λ are within the designer's control. Given that n has been reduced to the present day limit, and that b has been determined by the pulse length T required for a given range resolution (eqn. (17)), the designer is left only with P_t, A and λ as variables. Since the ratio A/λ enters as a square term, every effort must be made to produce a large aperture capable of working at the shortest possible wavelength. This creates severe engineering difficulties and the ratio of A to λ finally achieved is principally determined by the size of the budget. The largest steerable radio telescope so far built (at Jodrell Bank in Cheshire) has an effective aperture of 3000 square metres, at a wavelength of about 30 cms. At shorter wavelengths the irregularities in the surface of the reflector cause some defocussing of the beam and consequently a reduction of the effective aperture. The 600 ft. radio telescope being built by the U.S. Navy in Virginia will be capable of working to even shorter wavelengths (10–15 cms). To achieve this it is proposed to make the surface of separate rigid sections which can be adjusted to give the optimum point focus, whatever the position of the telescope.

The power P_t of the transmitter is another factor which is largely controlled by expense. For frequencies up to 1000 Mc/s it is quite possible to make transmitters having peak powers of several megawatts. For instance the transmitters currently used in early warning radar systems have output powers of the order of two megawatts and mean powers of one to two hundred kilowatts. Such transmitters represent what is currently possible, but they cost about $ 500 000.

Much of the work about to be described has been tackled with far less ambitious equipment. Thus the application of the most recent technical advances to the field of radar astronomy should produce a big reward. Nevertheless the number of objects which we shall be able to study in the near future seems to be confined to the sun and the nearer planets.

5. OBSERVATIONS OF THE MOON

The first group of workers to detect radio echoes from the moon were members of the U.S. Army Signals Laboratory, under the direction of Lieut.-Colonel J. H. De Witt. They succeeded in detecting echoes first at 11.58 a.m. on January 10 1946 (Webb, 1946; Mofenson 1946) but a full account of this work was not published until 1949 (De Witt and Stodola 1949). The parameters of their equipment were $P_t = 3$ KW (later raised to 15 KW); $f = 115{\cdot}5$ Mc/s, pulse length $T = 0{\cdot}2$ to $0{\cdot}5$ seconds; aerial gain $G = 24$ dB, receiver noise figure $n = 5$ dB, receiver bandwidth $b = 50$ cps. The aerial system took the form of an array mounted on the top of a 100 ft. tower which was rotatable so that observations could be made at moonrise and moonset. In their calculations De Witt and Stodola assumed that $g = 1$ and $\rho = 0{\cdot}17$, and hence a signal to noise ratio P_r/P_n of 1000 (+30 dB) was expected. The signals showed fading from pulse to pulse (the interpulse period was 4 seconds) and on some days were entirely absent. The maximum echo amplitudes observed were only in the region of +20 dB, and this discrepancy together with the fading could not be satisfactorily accounted for.

In 1949 workers at the Australian Commonwealth Scientific Research Organization (Kerr, Shain and Higgins 1949) commenced a series of observations using broadcast transmitters (VLC9 17·84 Mc/s, 50 KW output, or VLB5 21·54 Mc/s 70 KW output) and conventional communications-type receivers. The transmitter was keyed by landline signals from the receiving station some 600 Km distant. Kerr and Shain calculated that the post detector signal to noise ratio should be +25 dB for smooth moon scattering ($g=1$) and 32 dB for rough moon scattering ($g=5{\cdot}7$). A reflection coefficient $\rho=0{\cdot}15$ was assumed. The observed mean echo intensities frequently were about +10 dB and only occasionally rose to +20 dB above the noise level. At times the echoes completely disappeared. Despite the fact that the interpretation of their results was complicated by ionospheric effects (ionospheric refraction and possibly absorption) Kerr and Shain were able to distinguish two forms of fading. The first having a period of a few seconds was attributed to the moon's libration. This is the apparent rotation of the moon relative to an observer on the earth caused by (i) the observer's own east-west motion due to the rotation of the earth, (ii) longitudinal motion of the mean centre of the moon when viewed from the earth (caused by ellipiticity of the moon's orbit) and (iii) latitudinal motion of the mean centre caused by the fact that the axis of the moon's rotation is not perpendicular to the plane of its orbit. These three components of motion combine to give the moon an apparent rotation relative to an observer on the earth. This motion will depend upon the geographic latitude of the observer (being greatest for an observer at the earth's equator) and the rate of libration has both diurnal and monthly variations. Kerr and Shain (1951) were able to correlate the rate of fading of the echoes with the rate of librational spin. In this way they demonstrated that it is these changes in the aspect of the moon (causing changes in the phases of all the signals returned from the various scattering centres) which give rise to the fading.

The second fading mechanism appeared to be responsible for variations in the mean intensity of the echoes, and this had a period of the order of minutes. Kerr and Shain did not provide the proper explanation for this longer period fading, and the problem was not solved until Murray and Hargreaves (1954) discovered that it was produced by the rotation of the plane of polarization of the radio waves in the earth's ionosphere. This effect arises as a result of the presence of the earth's magnetic field, which causes the ionized medium to have two refractive indices. Thus plane waves transmitted along or nearly along the magnetic lines of force are propagated as two circularly polarized waves (of opposite sense) with different phase velocities. On reflection at the moon there is little depolarization, and hence when the two circular components are recombined back at the aerial they produce a plane wave which has been rotated relative to the original transmitted wave.

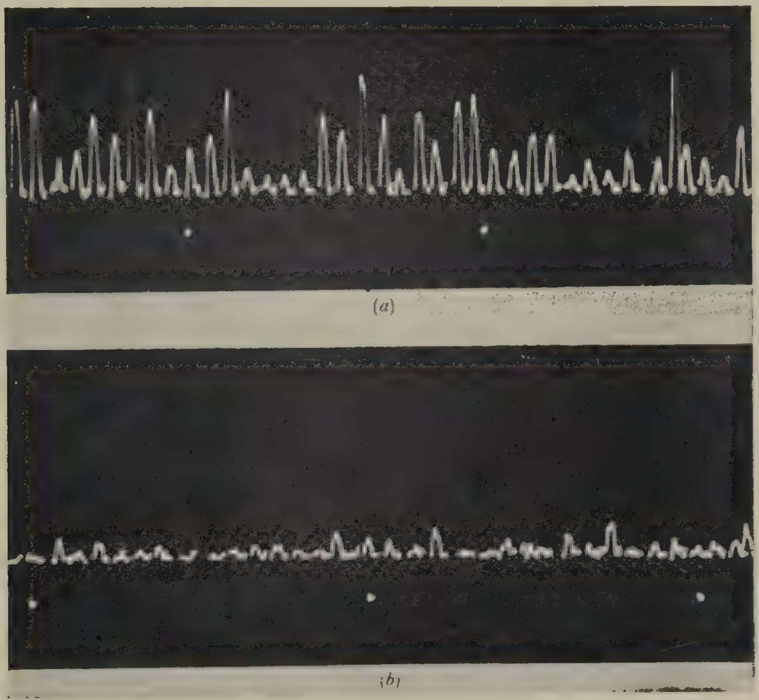
Browne *et al.* (1956) showed that the total amount of rotation is a function of the integrated number of electrons per cm^2 column through the ionosphere ($\int_0^R N$), and hence experiments to determine the total number of rotations of the plane of polarization of the radio waves would provide a new method of studying the ionosphere. Many such experiments have since been conducted for this purpose (Evans 1956, 1957 b; Bauer and Daniels 1958). Both signals reflected from the moon, and more recently observations of the radio signals from artificial earth satellites have been used. At a frequency of 100 Mc/s the amount of rotation will vary from as little as 1 turn at night to as many as 30 turns (for the two way journey) at midday. The number of rotations Ω can be determined by making use of the relation

$$\Omega \propto \frac{1}{f^2}. \qquad (18)$$

Hence a change $\delta\Omega$ in the total amount of rotation can be observed if the frequency is changed by an amount δf where

$$\frac{\delta\Omega}{\Omega}=\frac{2\delta f}{f}. \tag{19}$$

In this way Ω can be measured to give a value for the total number of electrons in the ionosphere. Such experiments permit the outermost parts of the earth's ionosphere, which are inaccessible to conventional ionospheric sounding equipments, to be studied for long periods of time. The two forms of fading, the rapid and the slow, are illustrated in fig. 5.


(a)

(b)

Fig. 5. Two samples of echoes observed by Browne *et al.* (1956). In each sample only the echoes have been photographed and these occur every 1·8 seconds. The dots below the traces are half-minute time marks. In both (*a*) and (*b*) the echoes are fading rapidly (due to the libration of the moon). The echoes in (*b*) which were observed some 14 minutes before those in (*a*) are on average much weaker because of the rotation of the plane of polarization in the earth's ionosphere.

The equipment used by Murray and Hargreaves (working at the Jodrell Bank Experimental Station) operated at a frequency of 120 Mc/s. Their aerial system took the form at a stepped broadside array of 10 sections having an aperture $A=250$ m^2 ($G=27$ dB). This had a beam directed due south and observations could be made for an hour every day for two weeks out of each lunar month. Evans (1957 a) used this equipment to study the brightness distribution across the surface of the moon. Because the doppler broadening introduced by the moon's libration is at the very most ±2 c/s. for observations at 120 Mc/s. Evans could not make this measurement using a narrow band receiver as a spectrum analyser. Instead he chose to measure the rate of the rapid fading caused by the moon's libration. This can be defined by the autocorrelation coefficient $\gamma(\tau)$ given by

$$\gamma(\tau)=\frac{\overline{y(t)y(t+\tau)}-\overline{y(t)}^{2}}{\overline{y^2(t)}\quad-\overline{y(t)}^{2}} \tag{20}$$

where $y(t)$ denotes the echo amplitude at a time t and $y\ (t+\tau)$ at a time τ later. The ' bars ' denote mean values. The autocorrelation function $\gamma(\tau)$ observed at any time will depend upon the rate of libration l_T at that time. This rate of libration l_T may be defined by the corresponding maximum doppler broadening of the echo f_o' since

$$f_o'=\pm\frac{2l_TR_o}{\lambda} \tag{21}$$

and hence the function $\gamma(f_o'\tau)$ should be time-independent and related to the power spectrum $P(f')df'$. A linear detector was used for which the autocorrelation function $\gamma(f_o'\tau)$ is given approximately by the square of the Fourier cosine transform of the radio frequency power spectrum $P(f')df'$. By sending out pairs of pulses each 20 milliseconds long and varying their separation τ Evans was able to determine $\gamma(f_o'\tau)$ and thence the power spectrum. The results of this work are shown in fig. 6 where the observed power spectrum is plotted together for comparison with that expected for Lommel-Seeliger and Lambert scattering. It is immediately evident that there is little doppler broadening of the signals. Some 50 per cent of the power is returned from a central region having a radius of 1/10th of that of the moon, and nearly all the power from within the first 1/3rd of the total radius. This rather surprising result was confirmed

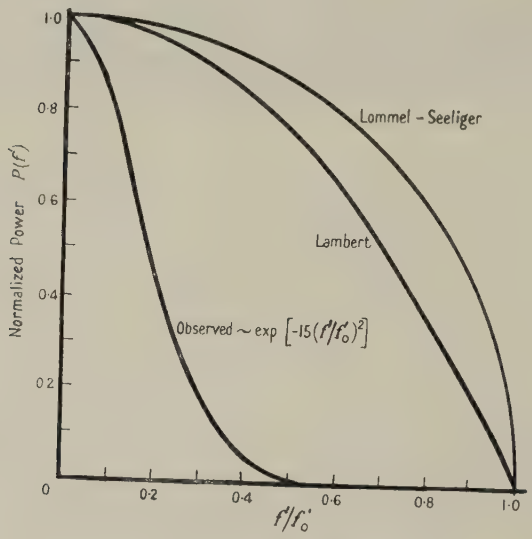


Fig. 6. The power spectrum of moon echoes observed by Evans (1957 a) compared with that to be expected if the Lommel-Seeliger or Lambert scattering laws were obeyed.

by observing the range and distortion of 2 milliseconds pulses. These appeared to be returned only from the front edge of the moon and showed no measurable elongation indicating that the depth of the effective scattering region was not greater than about 500 microseconds.

Evans postulated that since the Lambert scattering law

$$P_\phi = P_i \cos^2 \phi \tag{14}$$

gives rise to some limb darkening, then a law

$$P_\phi = P_i \cos^m \phi = \text{constant} \tag{22}$$

might be used to explain these results. By a process of numerical integration over the spherical surface the expected power spectrum $P(f')df'$ was calculated for several values of m. The observed and theoretical spectra were in close agreement for $m=30$. Thus a scattering law

$$P_\phi = P_i \cos^{30} \phi \tag{23}$$

was proposed, and the value of g was calculated by replacing $\cos\phi$ by $\cos^{30}\phi$ in the equations developed by Grieg, Metzger and Waer (1948), from which a value of $g=1{\cdot}8$ was obtained. From the measurements of the signal to noise ratio P_r/P_n which define the product ρg a value for

$$\rho = 0{\cdot}1 \pm\begin{matrix}0{\cdot}1\\ 0{\cdot}05\end{matrix}$$

was then calculated. Shortly afterwards Trexler (1958) published the results of some work carried out at the U.S. Naval Research Laboratories, Washington, several years previously. Trexler had tackled the problem of determining the distribution of the scattering centres by measuring the brightness variation with range instead of with doppler broadening. This required a high system sensitivity, and Trexler's equipment had the following parameters: $P_t=1$ M.W.; pulse length $T=12$ microseconds; frequency = 198 Mc/s; aerial gain = 40 dB; receiver noise figure $n=2$–3 dB; receiver bandwidth $b=250$ Kc/s. The aerial system was a section of a parabola of revolution having an elliptical opening 220 by 263 feet. It was formed by making a depression in the ground of the correct shape and surfacing this with a wire mesh. By an ingenious method of moving the primary feed system, observations could be made over a period of about 1 hour. The echoes obtained with this equipment, showed a large intial spike followed by an exponential tail. Some 50 per cent of the echo power was returned within the first 50 microseconds of the pulse (see fig. 7). This corresponds to a depth of five miles and a region of only 210 miles in diameter. This region is about 1/10th of the total radius of the moon and hence there is good agreement between the two experiments. The exponential tail appeared to extend some 500 microseconds beyond the leading edge of the echo, but this appeared to be a function of the signal to noise ratio, because strong echoes showed extensions out to 1000 microseconds.

It is clear from these experiments that radio waves of the order of a metre in length are reflected by the moon principally by a number of specular reflecting regions at the centre of the moon's disc. These reflections predominate over the reflections from other regions normal to the line-of-sight at greater distances. A similar result was reported by Yaplee (1958) and later by Hey and Hughes (1959) who made observations at a wavelength of 10 cms. These workers both used conventional microwave radar transmitters giving 2 Mw. peak power for 2 to

5 microsecond pulses, together with parabolic aerial systems of about 50 feet in diameter. By comparing the echo intensity with the inclination of the surface to the line-of-sight at different distances from the leading edge of the echo, Hey and Hughes concluded that 50 per cent of the echo power was returned from regions inclined at slopes of less than 5°. They argued that since the fading of the echo amplitude exceeds 100 per cent then the phase modulation introduced by the reflecting surface is likely to be greater than 2π radians. By arguments similar to those developed by Hewish (1951, 1952) they concluded that the value of 5° for the scattering angle would be consistent with a mean scale for the phase modulation at the lunar surface of about 1 metre. That is, in a horizontal distance over the surface of 1 metre the reflected wave electric field pattern will on average show a phase change of 2π radians.

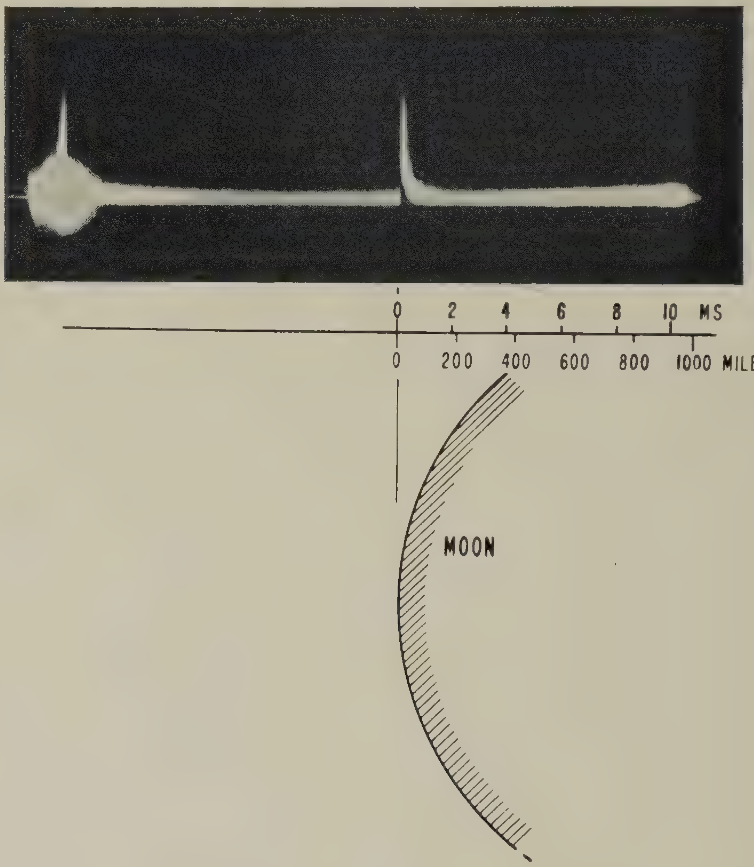


Fig. 7. The echoes observed by Trexler (1958), using short (12 microsecond) transmitter pulses. Most of the echo is returned from a very small region near to the front edge of the moon.

Hence the average surface gradient must be about 1 in 20. A similar result was obtained by Hargreaves (1959) who applied the theory developed for the study of the irregularities in the earth's ionosphere (see Ratcliffe, 1956) to the moon echo results obtained by Evans. He concluded that if the surface has height fluctuations about the mean with a standard deviation h, and if the correlation coefficient $\gamma(d)$ which relates these height variations to the horizontal distance over the surface follows a Gaussian law $\gamma(d) = \exp(-d^2/2d_0^2)$ then for the moon $h/d_0 \leqslant 0{\cdot}09$. Thus the gradients of regions of the surface one or more Fresnel zones in size (i.e. of the order of 1 kilometre across at radio wavelengths) are not greater than 1 in 10.

Since the moon is approximately uniformly bright at both light and infrared wavelengths (Markov 1958; Pettit and Nicholson 1930) it follows that the surface must be rough at wavelengths of the order of 10^{-2} cms. It is not known at what wavelength between $\lambda = 10$ cms and 10^{-2} cms the transition occurs between smooth moon scattering and uniformly bright scattering. It seems unlikely that this information could be obtained by earth-based radar equipment in view of the strong atmospheric absorption in this wavelength range.

Recently a more detailed examination of the scattering properties of the moon has been conducted by Pettengill (1960) who employed the Millstone Hill radar equipment operated by the Massachussetts Institute of Technology. His equipment parameters were; $P_t = 2{\cdot}1$ Mw.; pulse length $T = 500$ microseconds; frequency $f = 440$ Mc/s; receiver noise figure $n = 2$ dB and pulse repetition frequency $= 33{\cdot}4$ cps. An 84 foot parabolic aerial system was used for both transmitting and receiving. The local oscillator and transmitter frequencies used in this equipment were derived from a single stable oscillator, so that a chain of phase coherent pulses could be transmitted. These had an interpulse period of greater than the depth of the moon (11·6 milliseconds). The echoes were examined over 26 successive 500 microsecond range intervals, and the values for echo intensity and phase at each of the range intervals were stored by a fast digital computer for a period of operation of up to 10 seconds. Because a train of coherent pulses had been transmitted, Pettengill was able to obtain almost the same frequency resolution, that would have been observed for a C.W. signal lasting 10 seconds (i.e. $\pm 1/10$ cps). This was achieved by instructing the computer to perform a Fourier analysis of the stored amplitude and phase information to provide the power spectrum of the echoes for each of the 26 range intervals. The computer was made to present these on a cathode ray tube to give a display such as that shown in fig. 8. In this figure each horizontal line represents one of the 26 range intervals, and the abscissa is marked off in cycles of doppler broadening. The intensity of the echo power is presented in the form of histograms for each range interval. These have been scaled relative to each other by a factor which is shown in the left hand column, in order that they can all fit the display. These scaling factors show clearly that the most intense part of the echo arises as a result of reflections near the leading edge. However, the semi-circular outline of the histograms also shows that the echoes are produced at all ranges up to the limbs of the moon. The apparent limb brightening observed for the most distant range intervals is caused by the fact that whereas the contours of equal distance are circles concentric about the centre of the disc as seen from the earth, the contours of equal doppler shift are lines drawn parallel to the apparent axis of librational rotation. Thus near the limbs large areas fall inside both sets of contours, and give rise to strong high frequency components. Pettengill also used his computer to average the echo amplitudes as a function of range over a period of operation lasting 16 minutes. In this way an average $\sigma(\tau)$ function was obtained and is

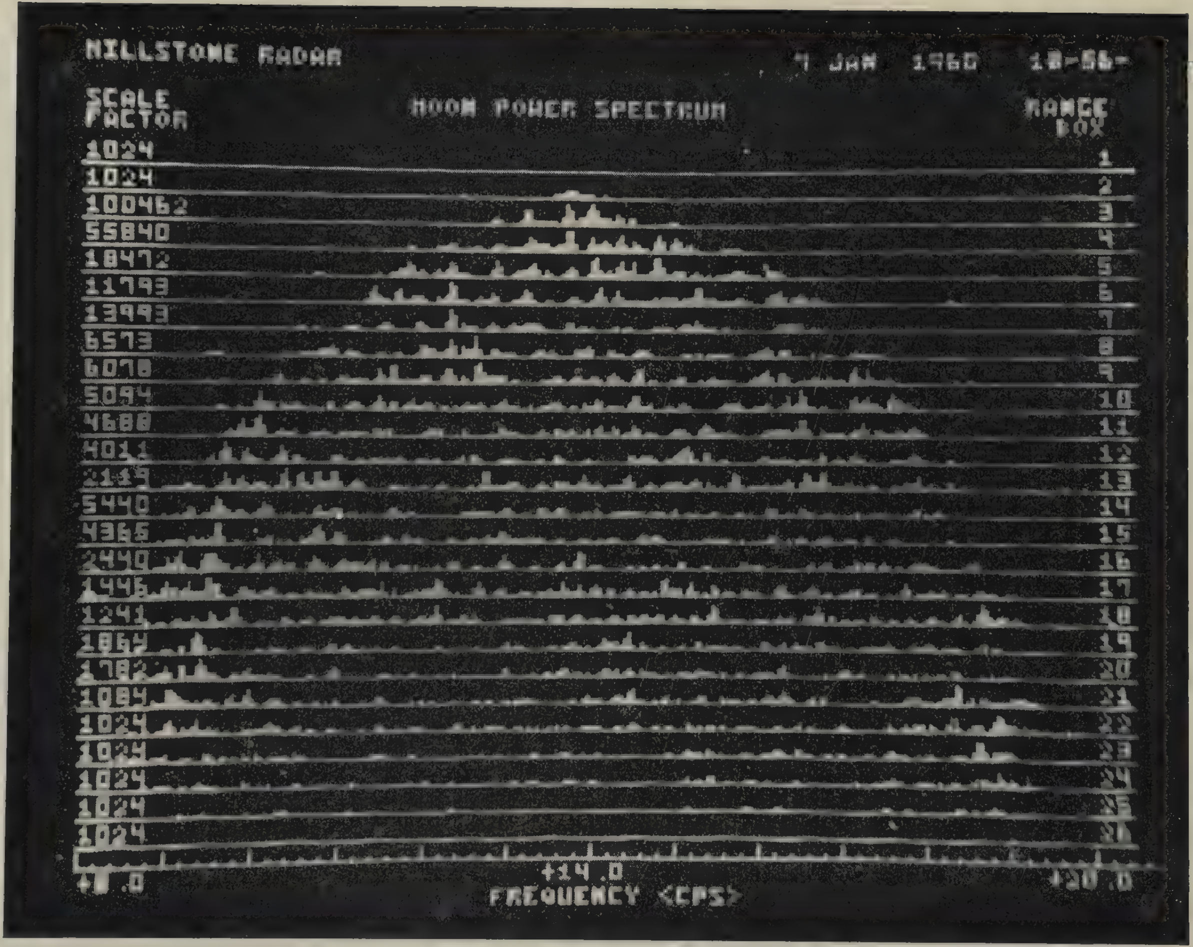


Fig. 8. A lunar 'map' at radio frequencies made by Pettengill (1960). In this map the echo intensity is plotted as a function of range and doppler broadening. The histograms have been scaled to one another by an amount shown in the left hand column.

plotted in fig. 9, as intensity against the cosine of the angle of incidence of the rays. As in the earlier work a strong initial echo having an exponential decay can be seen, but beyond this there is a weaker component which obeys the Lambert scattering law. The intensities of the two types of scattering become about the same at a radial distance from the centre of the moon's disc of about half the total radius. By comparing the relative amounts of specular and diffuse reflection Pettengill concludes that only about 10 per cent of the surface is sufficiently rough at this wavelength (70 cm) to give rise to Lambert scattering.

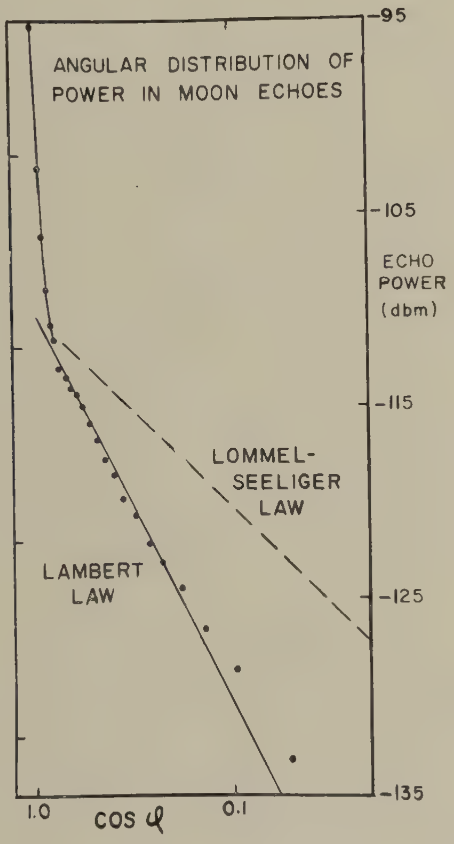


Fig. 9. The distribution of echo power as a function of the cosine of the angle of incidence of the radio waves to the surface observed by Pettengill (1960).

At the time of writing no visible features on the surface of the moon have yet been related to any of the observed echoes despite the great improvements in frequency and range resolution which have become possible. This is partly because any given set of range and frequency coordinates has two corresponding positions on the moon's surface. By observing over days where there are marked changes in the orientation of the axis of libration it may be possible to overcome this difficulty.

Many measurements have been made by other workers of the average echo intensity over the wavelength range 3·0 to 0·33 metres. If it is assumed that the scattering has no marked wavelength dependence over this range (none can be discovered from the measurements) a mean of the results can be taken and yields for the product ρg a value

$$\rho g = 7{\cdot}6 \pm 2{\cdot}1 \times 10^{-2}.$$

If it is further assumed that the value for g $(=1{\cdot}8)$ derived by Evans (1957a) is correct, then $\rho = 0{\cdot}042 \pm 0{\cdot}012$. By virtue of eqn. (12) (section 3·3) this corresponds to a value for the dielectric constant ϵ of $2{\cdot}3\,^{+0{\cdot}8}_{-0{\cdot}2}$ (another solution of the equation which gives a value of $\epsilon < 1$ has been ignored). It is difficult to reconcile this low value for the dielectric constant with that observed for commonly occurring terrestrial materials. For instance most silicates have dielectric constants of about 6, and fused quartz, one of the lowest, has a value $\epsilon = 3{\cdot}78$ (see " Reference Data for Radio Engineers ", 1953). This apparent disagreement may be resolved by postulating that the lunar surface materials are broken up and somewhat porous in texture.

Yaplee (1958) has made a study of the distance of the moon by making measurements at a wavelength of 10 cms with 2 microsecond transmitter pulses. The fading of the echo contributes the largest uncertainty to the measurements but by averaging over many echoes, the distance can be determined to within about half a mile. Marked differences between the observed and predicted ranges were observed at first, but by taking an improved value for the radial distance from the centre of the earth to Washington these discrepancies were reduced. The residual errors are now of the order of 5 miles, and these have yet to be accounted for. Yaplee hopes that by continuing his observations over a sufficient period of time he will be able to resolve the present discrepancies, and perhaps provide better values for the elements of the Earth-Moon system.

6. Observations of the planets

The relative scattering cross sections of the planets have been discussed earlier. Venus when at inferior conjunction has the next largest target term σ/R^4 after the moon, but the difference between the two is of the order of 10^7. Up to the present time there have been three serious attempts to detect echoes from Venus. The first attempt was made by workers at Lincoln Laboratory of M.I.T. (Price *et al.* 1959) who used the Millstone Hill Radar Station. The observations were made shortly after the close approach of Venus in February 1958, and the parameters of the equipment at that time were $P_t = 280$ Kw, $f = 440$ Mc/s, pulse length $T = 2$ milliseconds. An 84 ft. paraboloid was used for transmitting and receiving and a maser was used to give a low receiver noise factor. A train of pulses was transmitted for a period of five minutes, and then the receiver output was recorded for the following five minutes. No echoes greater than the noise level could be observed, or indeed were expected. Instead it was necessary to search in the recorded noise signals for a weak echo. This was done by means of a computer which stored the values for the noise amplitudes observed during the five minute period of reception, and performed a cross correlation process between these numbers and the pulse train waveform which would have been observed in the absence of noise. This pulse train is not exactly the same as the transmitted waveform since the range of Venus was increasing during the experiment, causing the echoes to be returned with a larger interpulse period than when transmitted. Such an integration process as this is an attempt to add all the echoes together to give a single echo greater than the noise, but provides no means of telling the exact range of the target. This will be uncertain by an amount equal to several times the interpulse period. To overcome this difficulty Price *et al.*, coded the transmitted waveform by missing out certain pulses from what would have otherwise been a continuous train of pulses. The same code pattern was later fed into the computer, which performed the cross correlation process missing out the same pulses, and in this way it was hoped to achieve an unambiguous range measurement.

Of the many records made only four proved suitable for complete analysis. On two of these a significantly high value of correlation was observed, and though the records were made on different days they agreed as to the range of Venus to within 1 millisecond. The two other runs showed no evidence of an echo. The value for the solar parallax derived from the experiment ($\pi = 0{\cdot}8022 \pm 0{\cdot}0001$ secs of arc) is plotted in fig. 10 together with the values for the earlier optical determinations. It can be seen that the optical results disagree by amounts for larger than their ‘ probable errors ’. The echo intensity was estimated as being 11 dB below the noise level, which indicated that the product ρg for Venus must be nearly unity. This seemed unduly large.

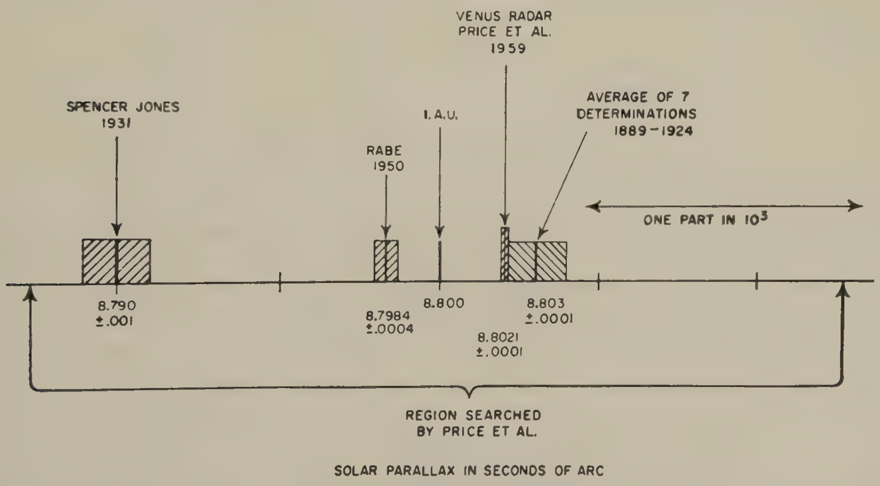


Fig. 10. The values for the solar parallax determined by optical and radar methods (diagram by Price).

In 1959 during the close approach of Venus in September the workers at Lincoln Laboratory repeated their experiment whilst Evans and Taylor (1959) attempted a similar experiment at Jodrell Bank. This work had the advantage of a larger radio telescope, but suffered because the transmitter power P_t was only 50 Kw. Thw other equipment parameters were: pulse length = 30 milliseconds, $f = 408$ Mc/s, receiver noise figure $n = 5$ dB receiver bandwidth $b = 60$ cps. Evans and Taylor also developed an integrating system to detect echoes weaker than the noise. In this system eight adjacent range intervals were examined to determine the average power level in each interval. The position of the intervals on the time-base could be adjusted so that the motion of the target could be compensated, and an echo if present would stay in the same interval. The range of Venus was computed in advance using the value of the solar parallax determined by Price *et al.* After some 60 hours useful observation one of the intervals showed a power level some $2\frac{1}{2}$ standard deviations higher than the mean of the others. If this were attributed to the planetary echo then the value of the solar parallax would be $\pi = 8{\cdot}8020 \pm 0{\cdot}0005$ secs of arc, in agreement with Price's determination, but the signal to noise ratio observed (in an integration system the signal to noise ratio increases as the square root of the number of echoes integrated) was some 18 dB lower than that predicted from Price's results. This suggested that Venus might scatter more like the moon with a value $\rho g \simeq 0{\cdot}05$. Meanwhile the workers at Lincoln Laboratory repeated their experiments using a higher transmitter power than previously, but this time without success. Only one recording made at Jodrell Bank of the output of a receiver tuned to the frequency of the American transmitter showed any evidence of an echo. Unfortunately from this recording the range could not be deduced. The failure of this later experiment to detect echoes throws doubt both on the earlier work, and on the Jodrell Bank observations, since the latter were based on the earlier radar determination of the solar parallax. It is possible that some fading mechanism is responsible for the anomaly, but this cannot be ionospheric rotation of the plane of polarization of the radio waves, as in both experiments circularly polarized waves were transmitted.

7. Observations of the sun

The sun is enveloped in a gaseous mantle known as the corona. In the corona the hydrogen gas is fully ionized and at a temperature which exceeds that of the visible surface of the sun—the photosphere. Much information concerning the sun's corona has been obtained both by optical methods during eclipses, and by radioastronomical studies.

Kerr (1952) and later Bass and Braude (1957) have discussed the possibility of observing radio echoes from the sun. Although at first sight it would seem than at echo from the sun would be some 100 times that from Venus (for a given equipment) there are two factors which make it difficult to detect solar echoes. These are: (i) The sun is an intense source of radio noise, and raises the effective aerial temperature T_A of a radio telescope pointed towards it by a large amount (see fig. 11). In addition to the so called ' quiet sun ' radio noise there are radio noise ' storms ' generated in the solar corona. These storms have been a subject of much study and are now classified into certain spectral types. Thus the minimum level of radio noise shown in fig. 11 is frequently exceeded by very large amounts (the frequency of these noise storms is roughly proportional to the sunspot number). (ii) Reflection will take place in the corona where the electron density N_e reaches a critical value appropriate to the frequency f (in megacycles/sec) of the exploring radio waves, where

$$N_e = 1{\cdot}24 \times 10^4 f^2 \qquad (24)$$

The higher the frequency, the greater is the depth to which the radio waves penetrate and as a result there is a corresponding increase in the attenuation due to absorption in the outer parts of the solar corona.

From a theoretical study of these effects both Kerr, and Bass and Braude considered that the optimum frequency to use for the detection of solar echoes would lie in the region 30–50 Mc/s. Kerr investigated the refraction and reflection process in the sun's corona for a frequency of 30 Mc/s. He concluded that because it seems likely that there are irregular variations of the density in the corona, the directivity term g will be of the order of 4, and the effective radius of the surface (due to the refraction effects) will be $1{\cdot}2\ R_0$. The average attenuation for the reflected rays at this frequency would be 3·5 dB (at 70 Mc/s the absorption would be increased by a further 10 dB). The doppler broadening was expected to be of the order of ±500 cps.

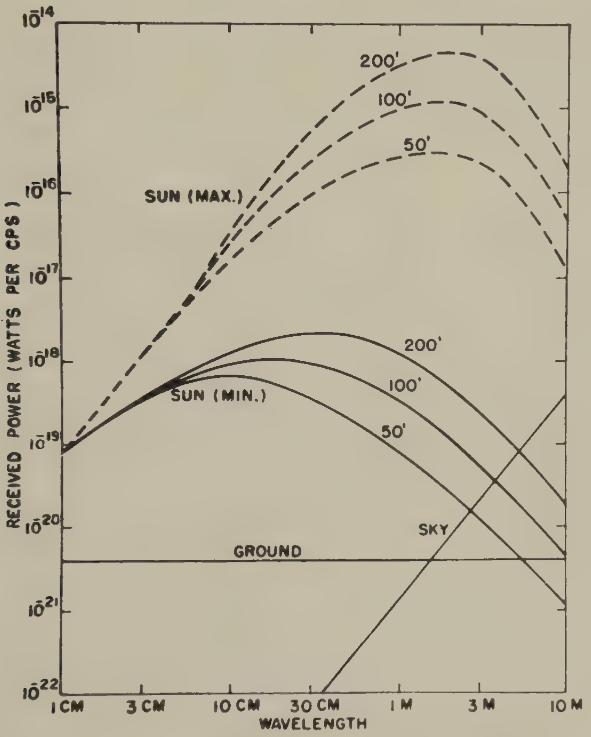
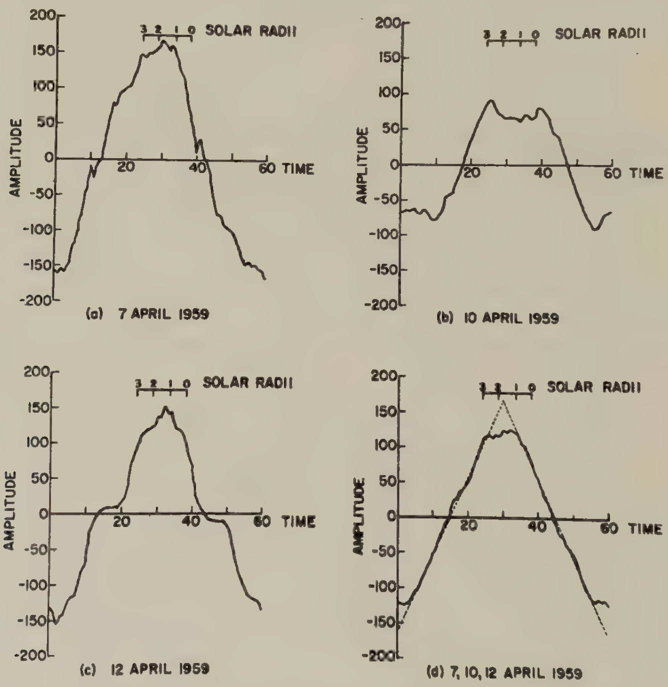


Fig. 11. The noise power received from the ground, the sky, and the sun for a parabolic reflector of 50, 100 or 200 ft. in diameter (diagram by Kraus).

In 1959, Eshleman *et al.* (1960) attempted to observe radio echoes from the sun. These workers (at Stanford University in California) employed a transmitter power $P_t = 40$ Kw, a frequency $f = 25{\cdot}6$ megacycles, an aerial gain $G =$ 25 dB, and a receiver bandwidth $b = 2$ kc/s. Their aerial system consisted of four rhombic aerials which gave a beam directed east at an elevation of 10°. The sun crosses this beam for 30 minutes shortly after sunrise on a few days each equinoctal period. The time delay for echoes from the sun is about 1000 sec, and Eshleman and his colleagues transmitted pulses each lasting 30 sec spaced by intervals of 30 sec for a total period of 900 sec. The aerial was then connected to the receiver and the transmitter turned off. The post-detector signal to noise ratio expected on the basis of Kerr's calculations was −44 dB, and thus as in the Venus echo attempts it was necessary to perform an integration process in order to detect the echo. The system employed was similar to that used by Price *et al.* The output of the receiver was recorded on magnetic tape and the amplitude of the noise signals subsequently measured. Next a cross correlation process was performed between these numbers and the expected echo waveform. The presence of an echo would cause a peak in the correlation curve corresponding to a given range. Three records obtained by the Stanford workers were sufficiently free from solar noise bursts to be worth analysing, and all three showed strong evidence of the echo (see fig. 12). The range of the echo inferred from fig. 12*d*, where the three results have been combined is 1·7 R_0 (R_0 is the sun's radius) from the centre of the sun. This is in reasonable agreement with the expected range and the observed signal to noise ratios were in remarkably good agreement with Kerr's predictions.



(a) 7 APRIL 1959

(b) 10 APRIL 1959

(c) 12 APRIL 1959

(d) 7, 10, 12 APRIL 1959

Fig. 12. The cross correlation diagrams obtained by Eshelman *et al.* (1960) which represent the integrated echo amplitude observed in their sun-echo experiments.

8. Future work

Apart from the sun and the planets, radar methods should be able to yield information concerning the comets and meteorites in interplanetary space. There is also the possibility of detecting the presence of ionized particles (either because they produce polarization rotation as in the ionosphere, or because they modify the refractive index and consequently the velocity of the radio waves). Most of this work depends upon the development of suitable radar systems. The U.S. Navy's 600 ft radio telescope when completed will be the largest fully steerable telescope, but as yet no details are available of the experiments which have been planned for it.

Cornell University is building a 1000 ft radio telescope. This will be excavated out of the ground (taking advantage of a suitable natural depression) in Puerto Rico. This reflector will be spherical rather than parabolic in section, and though this calls for a complicated primary feed system to give proper illumination it does permit the beam to be swung over a wide range of angles (up to about 50° from the zenith). Consequently moving objects can be followed for reasonable periods of time. It is proposed to equip this aerial system with a 400 Mc/s transmitter of the type already in use at Millstone Hill (2 MW peak power for 2 m.sec pulses). Although the principal purpose of the equipment will be to study the earth's ionosphere (the random scattering caused by the electrons in the ionosphere can be detected with a sensitive radar system) it is also intended to make planetary radar studies. Because the sensitivity of the system will be of the order of 40 dB greater than the Millstone Hill equipment, Venus, Mercury, Mars, Jupiter and possibly Saturn, may be within range. An alternative approach to obtaining a higher sensitivity has been adopted by M.I.T. who propose to build a telescope only 120 ft across, yet which can be used up to about 10,000 Mc/s.

To continue the sun echo programme Stanford University is installing a transmitter of 300 kw power, and is also building a 140 ft fully steerable radio telescope. The National Bureau of Standards is building a high power, low frequency radar system, which like that being built by Cornell University will be used principally for ionospheric scatter studies, but which will have applications to planetary radar. This equipment will work at a frequency of about 50 Mc/s with a transmitter power of several megawatts. The aerial system will be a linear array covering an area of four acres, and will be built near the geomagnetic equator in Peru. This equipment will certainly be able to detect echoes from the sun, which will be stronger than noise, and promises to reveal much about the structure of the solar corona. M.I.T. have plans to observe radio echoes from the sun using a transmitter of about 500 kw at a frequency of the order of 30 Mc/s and an aerial array some 200 ft by 1700 ft in aperture. This system will be sited at El Campo in Texas. The Jodrell Bank equipment is currently being improved for further attempts to detect echoes from Venus and possibly Mars.

Thus though little has been learned so far about the sun and the planets by radar methods, there should shortly be a considerable number of workers in the field equipped with suitable apparatus. We can but hope that the results provided by these new systems will justify their vast cost, and that perhaps a few more clues will be provided which will help unravel the age-old problem—the origin of the solar system.

REFERENCES

BASS, F. G., and BRAUDE, S. IA., 1957, *Ukr. Journal of Phys.*, **2,** No. 2, 149.
BAUER, S. J., and DANIELS, F. B., 1958, *J. Geophys. Res.*, **63,** 439.
BREIT, G., and TUVE, M. A., 1926, *Phys. Rev.*, **28,** 554.

Browne, I. C., Evans, J. V., Hargreaves, J. K., and Murray, W. A. S., 1956, *Proc. Phys. Soc. B.*, **69**, 901.

De Witt, J. H., and Stodola, E. K., 1949, *Proc. I.R.E.*, **37**, 229.

Eshelman, V. R., Barthle, R. C., and Gallagher, P. B., 1960, *Science*, **131**, 329.

Evans, J. V., 1956, *Proc. Phys. Soc. B.*, **69**, 953.
1957 (a), *ibid*, **70**, 1105.
(*b*), *J. Atmos. Terr. Phys.*, **11**, 259.

Evans, J. V., and Taylor, G. N., 1959, *Nature*, **184**, 1358.

Green, P., 1959, *privately communicated.*

Grieg, D. D., Metzger, S., and Waer, R., 1948, *Proc. I.R.E.*, **36**, 652.

Hargreaves, J. K., 1959, *Proc. Phys. Soc. B.*, **73**, 536.

Hewish, A., 1951, *Proc. Roy. Soc.*, **209**, 81.
1952, *ibid*,. **214**, 494.

Hey, J. S., and Hughes, V. A., 1959, *International Astronomical Symposium on Radio Astronomy*, pp. 13, 18 (Stanford: Stanford University Press).

Kerr, F. J., 1952, *Proc. I.R.E.*, **40**, 660.

Kerr, F. J., and Shain, C. A., 1951, *Proc. I.R.E.*, **39**, 230.

Kerr, F. J., Shain, C. A., and Higgins, C. S., 1949, *Nature*, **163**, 310.

Kraus, J. D., and Ko, H. C., 1957, *Electronics Research Directorate*, Air Force Cambridge Research Center, AFCRC-TN-57-557.

Lovell, A. C. B., 1954, *Meteor Astronomy*, pp. 51, 85 (Oxford, Clarendon Press).

Markov, A. V., 1948, *Astron. J., U.S.S.R.*, **25**, 172.

Mofensen, J., 1946, *Electronics*, **19**, 92.

Murray, W. A. S., and Hargreaves, J. K., 1954, *Nature*, **173**, 944.

Norton, K. A., and Omberg, A. C., 1947, *Proc. I.R.E.*, **35**, 4.

Petit, E., and Nicholson, S. B., 1930, *Astrophys. J.*, **71**, 102.

Pettengill, G., 1960, *Proc. I.R.E.* (in publication).

Price, R., Green, P. E., Goblick, T. J., Kingston, R. H., Kraft, L. G., Pettengill, G. H., Silver, R., and Smith, W. B., 1959, *Science*, **129**, 751.

Ratcliffe, J. A., 1956, *Rep. Prog. Phys.*, **19**, 188 (London: Physical Society).

Reference Data for Radio Engineers: Third Edition, 1955, pp. 48, 53 (New York: American Book—Stratford Press Inc.).

Senior, T. B. A., and Siegel, K. M., 1960, *Radio Propagation*, Journal of the National Bureau of Standards (in publication).

Smith, M. A., 1951, *Proc. I.E.E.* **98**, Part IV.

Trexler, J. H., 1958, *Proc. I.R.E.*, **46**, 286.

Webb, D. H., 1946, *Sky and Telescope*, **5**, 3.

Winter, D. F., 1956, *Electronic Research Directorate*, Air Force Cambridge Research Center, AFCRC-TR-56-106.

Yaplee, B. S., Bruton, R. H., Craig, K. J., and Roman, N. G., 1958, *Proc. I.R.E.*, **46**, 293.

See also

Bay, Z., 1946, *Hung. Phys. Acta*, **1**, 1.

Blevis, B. C., 1957, *Nature*, **180**, 138.

Hanbury Brown, R., and Lovell, A. C. B., 1957, *The Exploration of Space by Radio*, pp. 178, 190 (London: Chapman and Hall).

Pawsey, J. L., and Bracewell, R. N., 1955, *Radio Astronomy*, pp. 293, 305 (Oxford: Clarendon Press).

Sulzer, P. G., Montgomery, G. F., and Gerks, I. H., 1952, *Proc. I.R.E.*, **40**, 361.

Thomas, A. B., 1949, *Aust. J. Science*, **1**, 1187.

The Author:

The author graduated at Manchester University in 1954 and went to work at the Jodrell Bank Experimental Station. During the following five years he made radar studies of the reflecting properties of the moon's surface, made the first accurate measurements of the number of electrons in the earth's ionosphere, and designed a radar equipment with which echoes from Venus were detected in 1959. In 1960 he took a position at the Lincoln Laboratory operated by Massachusetts Institute of Technology. This article was contributed while he was at Jodrell Bank.

Non-linear Vibrations

by E. W. LEE
University of Sheffield

1. INTRODUCTION

We are all familiar with the idea of simple harmonic motion. If a body of mass m moves along a line such that the force on the body is always directed towards a fixed point and is proportional to the distance from that point, the equation of motion of the body is, by Newton's second law

$$m\frac{d^2x}{dt^2}+\alpha x=0 \tag{1}$$

The solution is

$$x=A\cos\omega_0 t+B\sin\omega_0 t \tag{2}$$

with $\omega_0{}^2=\alpha/m$. An equivalent form of eqn. (2) is $x=C\cos(\omega t-\phi)$. The constants A and B (or C and ϕ) which do not concern us greatly, depend upon the manner in which the body was set in motion. The subsequent movement is given by eqn. (2). Since

$$\cos\omega t=\cos(\omega t+2\pi)=\cos\omega(t+\frac{2\pi}{\omega})$$

it follows that the motion repeats itself *exactly* after a time $T=2\pi/\omega$. This time is called the period of the motion, the motion itself being described as simple harmonic. The frequency $\nu=1/T$. The period depends only on the mass of the body and the restoring force constant α, and is therefore independent of the amplitude. Such motion is described as isochronous.

Now, for a variety of reasons, true simple harmonic vibrations do not occur in nature. In the first place some damping is always present. As a result, the energy of the vibration is continuously dissipated and the amplitude (C) of the motion decreases. If the damping is viscous (i.e. the damping force is proportional to the velocity) then the equation of motion is

$$m\frac{d^2x}{dt^2}+2k\frac{dx}{dt}+\alpha x=0. \tag{1a}$$

which has as its solution

$$x=De^{-kt}\cos(\omega t-\phi) \tag{3}$$

where D is a constant. In this case ω is not equal to $(\alpha/m)^{1/2}$ but is given by $\omega^2=\omega_0{}^2-k^2/m^2$. The effect of damping is thus to decrease the frequency slightly. Moreover as a result of the exponential term in eqn. (3) the motion never repeats itself identically and is therefore not strictly simple harmonic. It is however isochronous† and so the free undamped period can always be obtained by measuring k and making appropriate corrections which, in any case, can be made small by appropriate adjustment of physical conditions.

---

† It will not remain so however if the amplitude and velocity are so great as to give rise to turbulence, for then the damping constant k will vary.

Secondly, by far the greatest number of vibrating systems do not obey Hooke's Law exactly, i.e. the restoring force is not in general proportional to the displacement. Consider, for example, the familiar case of a bar magnet freely suspended in a uniform field. If the axis of a magnet of a moment M is inclined at an angle ϕ to the field H, the restoring couple tending to make M and H parallel is $-HM \sin \phi$. The equation of motion is therefore

$$I\frac{d^2\theta}{dt^2}+MH\sin\theta=0. \tag{4}$$

The restoring couple is not proportional to the angular displacement and we cannot expect the motion of the magnet to be simple harmonic. Now,

$$\sin\theta=\theta-\frac{\theta^3}{3!}+\frac{\theta^5}{5!}-\ \ldots.$$

so that if the angle θ is not too large we may re-write eqn. (4) as

$$I\frac{d^2\theta}{dt^2}+MH\theta-\frac{MH\theta^3}{6}=0. \tag{5}$$

The third term can always be made negligible in comparison with the second by restricting the motion to very small (strictly, infinitesimal) angular amplitudes. In such circumstances the form of eqn. (1) is regained and the resulting motion is simple harmonic.

Solutions of eqns. (4) and (5) for finite amplitudes show the motion to be strictly periodic since damping is assumed to be absent, and almost simple harmonic but not isochronous.

Almost all the more familiar vibrating systems belong to this class exhibiting 'conditionally simple harmonic motion'. Obvious examples in addition to the vibrating magnet are the simple and compound pendulum, torsion pendulum, and bifilar suspension. Also, sound waves of finite amplitude come into this category since the characteristic relation used in developing the theory of sound propagation is the adiabatic, $pV^{\gamma}=\text{constant}$, and the graphical relation between p and V is a curve, no finite portion of which is linear. One final example is that of the stretched string. If a string, fixed at both ends is made to vibrate transversely the string must extend somewhat during the period of vibration in order to be able to vibrate at all. This gives rise to an extra tension in the string proportional to the cube of the transverse displacement so that the equation of motion is of the form of eqn. (5).

One of the chief characteristics of the vibrations mentioned so far is that they all revert to the simple harmonic type when the amplitude is restricted to very small values. However other types of vibrations occur which are not even approximately simple harmonic. Examples are the 'chattering' of friction brakes, the bumping in cold water pipes due to a worn tap washer, the squeaking of a dinner fork on a plate or of chalk on a blackboard, and the beating of the heart. It is a matter of common experience that oscillations of this kind frequently build up to very large amplitudes; under these conditions they are often known as relaxation oscillations. Electrical oscillations of this type are produced by the multivibrator and blocking oscillator.

All these oscillating systems have one thing in common. Their differential equations of motion are non-linear, that is to say, they contain the dependent variable y and/or its differential coefficients raised to powers higher than the first degree. The physical origin of the non-linearity may be different. The vibrating magnet and the pendulum are governed by *non-linear restoring couples or forces*, whereas relaxation oscillations arise from *non-linear* (*negative*) *damping terms*. Only a small number of non-linear differential equations admit an exact solution and frequently one has to resort to approximate analytical methods. We shall discuss the behaviour of a few non-linear systems by these means.

2. Systems having non-linear restoring force

2.1. *Free vibrations*

These are among the conditionally simple harmonic systems discussed above. We consider first the case of the vibrating magnet. The equation of motion is

$$I\frac{d^2\theta}{dt^2}+MH\sin\theta=0 \tag{6}$$

This equation may be solved exactly as follows. Put $d\theta/dt=v$, $MH/I=a$. Then eqn. (6) becomes

$$v\frac{dv}{d\theta}+a\sin\theta=0$$

so that

$$\int v\,dv=-a\int\sin\theta\,d\theta+\text{ const.}$$

$$\therefore\ v^2=2a\cos\theta+\text{ const.}$$

If the amplitude of the swing is θ_0 then, since this occurs when $v=0$, we have

$$-2a\cos\theta_0=\text{const.}$$

and so

$$v=\pm(2a)^{1/2}(\cos\theta-\cos\theta_0)^{1/2}$$

$$=\pm 2a^{1/2}\left(\sin^2\frac{\theta_0}{2}-\sin^2\frac{\theta}{2}\right)^{1/2}$$

If we now introduce two new variables k and ϕ, defined by $k=\sin\theta_0/2$, $k\sin\phi=\sin\theta/2$, equation becomes

$$v=\frac{d\theta}{dt}=-2a^{1/2}k\cos\phi \tag{7}$$

the negative sign being chosen since v increases with decreasing θ. Now from the definition of ϕ we have

$$\tfrac{1}{2}\cos\frac{\theta}{2}d\theta=k\cos\phi\,d\phi$$

so that

$$d\theta=\frac{2k\cos\phi\,d\phi}{(1-k^2\sin^2\phi)^{1/2}}. \tag{8}$$

Elimination of $d\theta$ from eqns. (7) and (8) gives

$$dt=-\frac{d\phi}{a^{1/2}(1-k^2\sin^2\phi)^{1/2}}. \tag{9}$$

This equation, on integration, enables us to relate the angular displacement to the time. In particular we may calculate the periodic time in the following manner. If we take the instant $t=0$ to be at time of maximum displacement then $v=0$ and $\phi=\pi/2$. At $t=T/4$ we then have $\theta=0$, and $\phi=0$ so that

$$T/4=\int_0^{T/4} dt=a^{-1/2}\int_0^{\pi/2}\frac{d\phi}{(1-k^2\sin^2\phi)^{1/2}}. \tag{10}$$

The integral on the right is known as a complete elliptic integral of the first kind, modulus $k=\sin\theta_0/2$ denoted by $F(k, \pi/2)$. Hence the periodic time is

$$T=4\sqrt{\left(\frac{I}{MH}\right)}\cdot F(k, \pi/2). \tag{11}$$

From a table of elliptic integrals we find that when $k=0$, $F(k, \pi/2)=\pi/2$ so eqn. (11) reverts to the familiar form. The tables also show that F increases with increasing k, i.e. with increasing amplitude. By changing the limits of integration in eqn. (11) and using the tables of elliptic integrals it is possible to calculate the displacement as a function of time. The interesting feature which then emerges is that the motion remains very nearly simple harmonic even at amplitudes sufficiently great to cause a significant deviation from the periodic time for small amplitudes. In other words the non-linearity in the restoring force, although it destroys the isochronous properties of the vibrating system, does not sensibly affect the simple harmonic character of the motion. This is a very important result because it means that for vibrating systems of this kind where the equation of motion cannot be solved exactly, one may always assume an approximate solution of the form $x=a\cos\omega t$ with the assurance that the approximation is a good one.

To illustrate the procedure consider a vibrating system whose equation of motion is of the form

$$\frac{d^2x}{dt^2}+\alpha x+\beta x^3=0 \tag{12}$$

which is of the same form as the second approximation to eqn. (4) if $\beta=-\alpha/6$. Although eqn. (12) can be solved exactly it is instructive to use the method of successive approximations. In view of the statements made in the previous paragraph it is reasonable to suppose that an approximate solution is $x=a\cos\omega t$; (this is not the same ω as in eqn. (3)). If we substitute this in eqn. (12) and make use of the fact that $\cos^3\theta=\frac{1}{4}\cos 3\theta+\frac{3}{4}\cos\theta$ we obtain

$$-a\omega^2\cos\omega t+\alpha a\cos\omega t+\tfrac{1}{4}\beta a^3\cos 3\omega t+\tfrac{3}{4}\beta a^3\cos\omega t=0.$$

If we now ignore the third term, then $a\cos\omega t$ appears as a common factor and may be cancelled giving

$$\begin{aligned}\omega^2&=\alpha+\tfrac{3}{4}\beta a^2\\ &=\omega_0{}^2+\tfrac{3}{4}\beta a^2\end{aligned} \tag{13}$$

which expresses the variation of frequency with amplitude a. If we next substitute for αx and βx^3 in eqn. (12) and use eqn. (13) we obtain

$$\frac{d^2x}{dt^2}=-a\omega^2\cos\omega t-\frac{1}{4}\beta a^3\cos 3\omega t.$$

Integration gives

$$\frac{dx}{dt}=-a\omega\sin\omega t-\frac{1}{12}\frac{\beta a^3}{\omega}\sin 3\omega t+\text{const.}$$

The constant must be taken to be zero to avoid a term ct after further integration, which then gives

$$x = a \cos \omega t - \frac{1}{36} \frac{\beta a^3}{\omega^2} \cos 3\omega t. \tag{14}$$

Using eqn. (13) ω may be eliminated from the denominator of the second term which is then seen to depend only upon the quotent $\beta a^3/\alpha$. Equation (14) may be regarded as a second approximation to be substituted into the original equation to give a third approximation and so on. From eqn. (13) it can be seen that the frequency increases or decreases with amplitude according to whether β is positive or negative. If β is positive the restoring force is greater than that given by Hooke's Law and the system is said to be hard. If β is negative the system is described as soft.

If the system is not symmetric but contains even powers of x in the expression for the restoring force, i.e. $\alpha x + \gamma x^2$, then a similar procedure shows that, even harmonics are present. In addition there is a constant term representing unidirectional displacement of the mass point from the origin. This is the formal explanation of the thermal expansion of solids. The force-displacement relation for atoms in a solid is not only non-linear but asymmetric. If Hooke's Law were obeyed there would be no thermal expansion. The existence of non-linear terms in the force-displacement relation forms the basis of a theory of thermal expansion by Grüneisen which is in good agreement with experiment.

2.2. *Forced vibrations*

It must not be thought that a system which is described by eqn. (12) is merely a simple harmonic system perturbed slightly by the term in x^3. The existence of the non-linear term alters the nature of the system to such an extent that non-linear systems exhibit characteristics not possessed by simple linear systems. We see this most readily if we enquire into the response of such a system to a periodic driving force. The equation of motion is

$$\frac{d^2x}{dt^2} + \alpha x + \beta x^3 = F \cos pt. \tag{15}$$

We are now in a position to appreciate the success of the approximational method used above for free vibrations for it is a remarkable thing that this innocent looking equation has so far resisted all attempts to solve it explicitly. From our knowledge of the response of a linear system to a periodic force we may assume a first solution of the form $x = a \cos pt$. Proceeding exactly as before we find

$$-ap^2 \cos pt + \alpha a \cos pt + \tfrac{1}{4}\beta a^3 \cos 3pt + \tfrac{3}{4}\beta a^3 \cos pt = F \cos pt.$$

Ignoring the term in $3pt$ as before and eliminating $a \cos pt$ as a constant factor we obtain

$$(\omega_0{}^2 - p^2)a + \tfrac{3}{4}\beta a^3 = F. \tag{16}$$

Proceeding exactly as for the free vibrations, this time making use of eqn. (16). we obtain

$$x = a \cos pt + \frac{\beta a^3}{36p^2} \cos 3pt \tag{17}$$

as the second approximation. However the interesting thing to notice is eqn. (16) which expresses the amplitude as a function of the driving frequency p, for this is a cubic in a and this implies that over a certain frequency range three different amplitudes are possible. To investigate this more closely we solve eqn. (16) for the amplitude. The simplest way to do this is to plot the cubic $y=\frac{3}{4}\beta a^3$ and the straight line $y=(p^2-\omega_0^2)a+F$ on a graph and find the points of intersection (fig. 1). The straight line has intercept F and slope $p^2-\omega_0^2$ so that for $p<\omega_0$ its slope is negative and there is only one point of intersection. As p increases however the line pivots about the point $y=F$ and ventually becomes tangential to the cubic in the negative quadrant and intersects it in the positive one so that two amplitudes are possible. As p increases still further there are three points of intersection corresponding to three possible amplitudes. The amplitude-frequency curves calculated in this way are shown in fig. 2. For positive values of β the resonance curve is bent over towards the high frequency side and in the absence of damping the amplitude would increase indefinitely with increasing frequency. In practice some damping is always present and this has the effect of uniting the two branches of the curve as shown by the thick line. Under these circumstances the amplitude grows with increasing frequency until the point A is reached where the tangent to the curve is vertical. This is a position of instability and if the frequency is increased further the amplitude drops suddenly to the stable state represented by B.

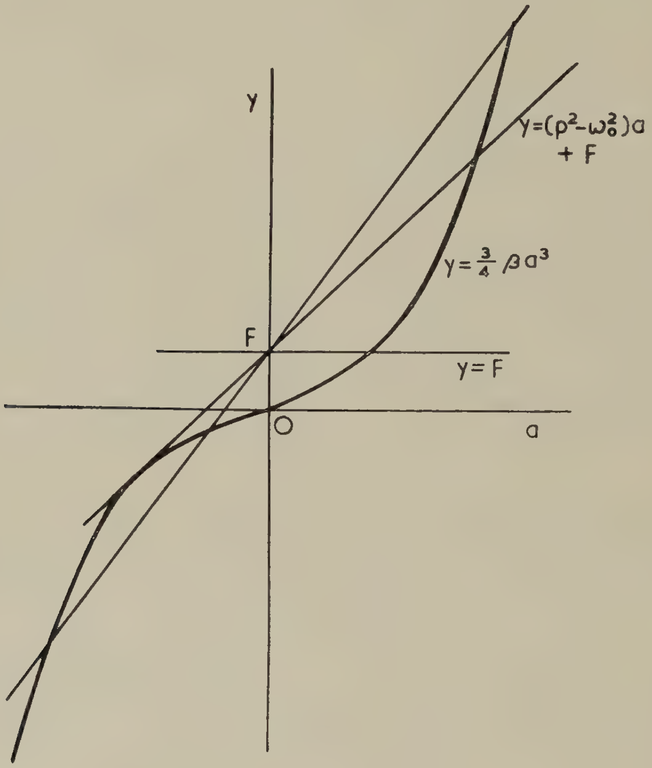


Fig. 1. Graphical solution of the equation $(\omega_0^2-p^2)a+\frac{3}{4}\beta a^3=F$.

Further increase of frequency causes the system to move outwards towards the region C, the amplitude decreasing slowly. If now the frequency is decreased the lower branch of the curve remains the stable one until the point D is reached and any further decrease in frequency is accompanied by a sudden jump in the amplitude to the point E on the upper branch. The system thus exhibits hysteresis in so far that the amplitude of the motion is not, unlike a linear system, a unique function of frequency but depends upon whether the operating frequency has been reached from a higher or a lower frequency. It may readily be verified that if β is negative a similar effect occurs but in this case the resonance curve is sheared over to the lower frequency side.

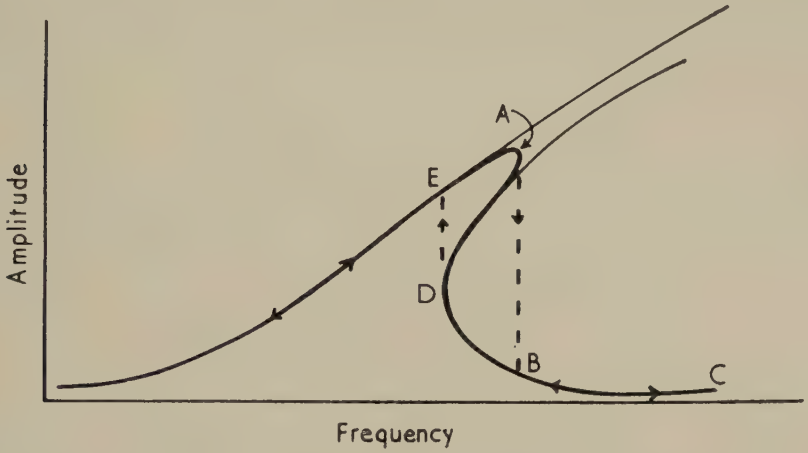


Fig. 2. Amplitude-frequency curves for non-linear forced vibrations;
——————— no damping;
– – – – – – with damping.

2.3. *A simple demonstration experiment*

A simple system which exhibits all the features encountered above is a stretched string driven transversely at its mid-point. The simple theory of the stretched string requires the tension in the string to be constant. Actually any transverse displacement must necessarily be accompanied by a slight increase in the length of the string, which if clamped at both ends therefore experiences an additional tension which to a first approximation varies as the cube of the displacement. A diagram of a suitable arrangement (almost identical with that sometimes used as a laboratory experiment to find the frequency of the A.C. mains) is shown in fig. 3. A Eureka (constantan) wire *clamped at both ends* is

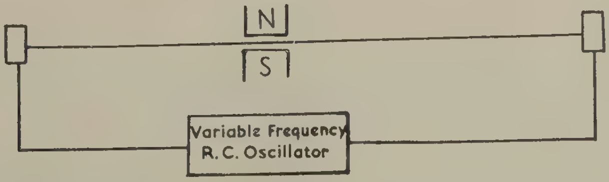


Fig. 3. Simple apparatus for observing forced vibrations of a non-linear system. It is important that the wire be clamped at both ends.

connected to an R-C oscillator. A permanent magnet (an old magnetron magnet serves well) placed at the mid-point of the wire produces a field perpendicular to the wire, which when carrying the alternating current from the oscillator experiences a transverse sinusoidal driving force perpendicular to the magnetic field. A typical resonance curve, exhibiting the characteristic 'jump phenomena' is shown in fig. 4.

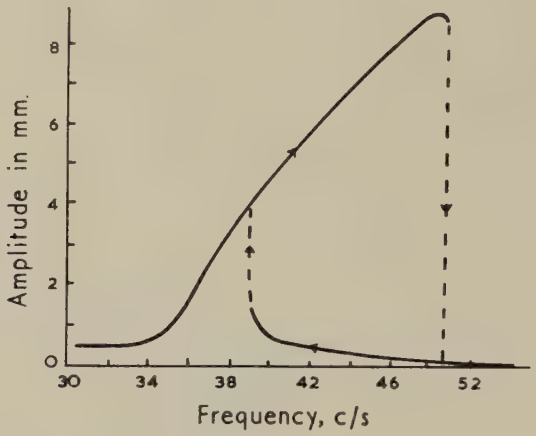


Fig. 4. Representative amplitude-frequency curve for the string of fig. 3.

2.4. *Subharmonic response*

Another characteristic of non-linear systems which is not found in linear systems is that they may exhibit subharmonic response. By this we mean that a system driven by a sinusoidal driving force of frequency p responds by vibrating with a frequency p/m where m is an integer. The permitted values of m depend on the type of non-linearity which is present and for a system obeying the equation

$$\frac{d^2x}{dt^2}+\alpha x+\beta x^3=F\cos pt. \tag{18}$$

m is equal to $\frac{1}{3}$ so that a solution of the form $x=a\cos\frac{1}{3}pt$ exists. If the third term in eqn. (18) is quadratic instead of cubic, $m=2$ and a vibration $x=a\cos\frac{1}{2}pt$ may occur. There are grounds for believing that if the restoring force is of the form $\alpha x+\beta x^3+\gamma x^5+\ldots$ then subharmonics of order 1/3, 1/5 and so on will be observed.

The conditions under which subharmonic terms may occur are derived in the books by Stoker and McLachlan. To prove that subharmonic response does occur is an altogether different problem which, as far as we are aware, has not yet been satisfactorily solved. Fortunately subharmonic oscillations have been observed in certain non-linear systems and indeed it is this very characteristic which makes the extremely non-linear valve oscillator known as the multivibrator so useful as a frequency divider.

2.5. *Intermodulation and Combination Tones*

A linear vibrating system obeys the principle of superposition. If acted upon by two sinusoidal forces of frequency p_1 and p_2 its response may be obtained by addition of two expressions each representing its response to the frequencies p_1 and p_2 separately. This principle is widely used in the mathematical treatment of optical interference. The situation is quite different for non-linear systems. If we take the equation of motion to be

$$\frac{d^2x}{dt^2}+\alpha x+\beta x^3=H_1\cos p_1t+H_2\cos p_2t \tag{19}$$

then we can assume a first approximate solution $x=A\cos p_1t+B\cos p_2t$ which is the form of solution if $\beta=0$. It may then be shown that the next approximation is

$$\begin{aligned}x=A\cos p_1t+B\cos p_2t+\frac{\beta A^3}{36}\cos 3p_1t+\frac{\beta B^3}{36}\cos 3p_2t\\ +\tfrac{3}{4}\beta A^2B\left\{\frac{\cos(p_2+2p_1)t}{(p_2+2p_1)^2}+\frac{\cos(p_2-2p_1)t}{(p_2-2p_1)^2}\right\}\\ +\tfrac{3}{4}\beta AB^2\left\{\frac{\cos(p_1+2p_2)t}{(p_1+2p_2)^2}+\frac{\cos(p_1-2p_2)t}{(p_1-2p_2)^2}\right\}\end{aligned}$$

in which A and B must satisfy the relations

$$(\alpha-p_1{}^2)A+\tfrac{3}{4}\beta A(A^2+2B^2)=H_1$$

$$(\alpha-p_2{}^2)B+\tfrac{3}{4}\beta B(B^2+2A^2)=H_2.$$

The point of interest here is that the non-linear term gives rise not only to the third harmonic terms of frequency $3p_1$ and $3p_2$ but also to components with frequencies $p_2\pm 2p_1$ and $p_1\pm 2p_2$, the so-called intermodulation frequencies, or, in acoustics, combination tones. If instead of the term $-\beta x^3$ in the left-hand side of eqn. (19) we had taken a term $-\gamma x^2$, a similar result would have been obtained except that combination tones consist of a summation tone of frequency p_1+p_2 and a difference tone of frequency p_1-p_2.

These facts are of considerable importance in the theory of hearing. Two features are noteworthy. The first is that tones may be present in the sensation to which there corresponds no constituent simple harmonic vibration in the external stimulus. Fletcher (1953) has shown that if a pure tone (i.e. simple harmonic vibration) of frequency is injected into the ear, sensations corresponding to harmonics of frequency $2p$, $3p$, $4p$ and so on are heard. Helmholtz gave reasons for believing that the tympanic membrane (ear drum) and the ossicles of the middle ear form a non-linear vibrating system and this has been confirmed by careful observations on a scale model of the ossicles. The system behaves rather like a soft spring ($\beta<0$), but is also asymmetric with respect to the origin. We can thus represent the restoring force by an expression of the form $\alpha x+\gamma x^2+\beta x^3+\dots$ and this explains in principle the presence of sensations corresponding to both even and odd harmonics.

If two intense pure tones are heard together a third tone can be heard whose frequency is the difference of the two generating tones. The discovery of this difference tone is usually ascribed to the Italian violinist Tartini. The once widely held view that such tones are the result of beats between the two frequencies was rejected by Helmholtz on the grounds that whereas beats are heard with feeble tones, fairly intense sound sources have to be used in order to hear these difference tones. He favoured the view that the difference tone arises from the non-linear behaviour of the ear and this led him to the discovery of summation tones. Of course one must always reckon with the fact that combination tones may be produced in the air itself since the adiabatics for air are curved, but the non-linear response of the ear seems now well established and explains in a natural way both the existence of aural harmonics and combination tones. An interesting brief account of this topic is given by A. Wood (1940) and a more detailed one in the book by Fletcher already cited.

3. Self-sustained oscillations

In the previous section we dealt with systems in which the non-linear element was provided by the restoring force. However, non-linearities in the other force terms may occur as we can see from the following general considerations. If in eqn. (1a) the damping constant happens to be negative then eqn. (3) indicates that the amplitude of the motion increases exponentially with time. However some positive damping at large amplitudes must occur if the vibrating system is to remain physically stable otherwise it would eventually break apart. A damping term of the form

$$-k\frac{dx}{dt}+k'\left(\frac{dx}{dt}\right)^3$$

will ensure that the damping is such as to increase the amplitude for small velocities but to decrease it for large velocities. It follows that the stable state is not one in which the system is at rest but is such that a steady oscillation is maintained in which the energy abstracted by the system due to the negative damping term is equal to that dissipated in it due to the positive one.

Negative damping occurs most naturally in systems subject to dry frictional forces. One of the best known examples is shown in fig. 5. Here a mass m

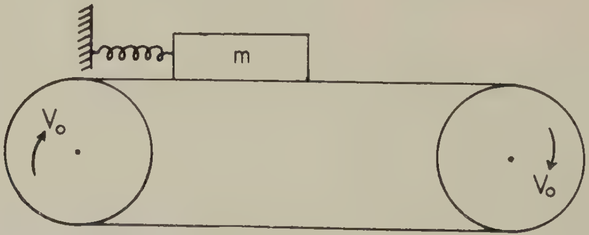


Fig. 5. Mechanical system capable of executing self-sustained oscillations.

connected to a spring s rests on a rough endless belt driven at a uniform speed v_0. The force on the mass is the restoring force $-\alpha x$ due to the spring, and the frictional force between mass and belt. This force depends upon the relative velocity between mass and belt in a rather complicated manner (fig. 6) ; at zero relative velocity it is represented by OA. As the relative velocity increases the frictional force decreases to minimum and increases thereafter. From

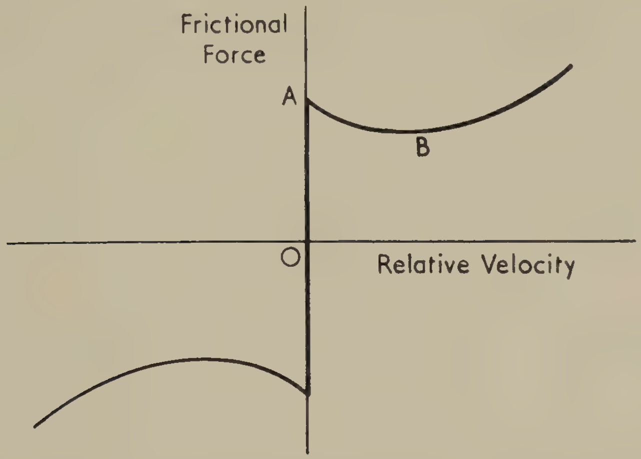


Fig. 6. Graph showing the form of relationship between solid friction force and relative velocity.

A to B the slope of the curve is negative and this implies that the damping constant is negative. If we denote the frictional force by $f[v_0-(dx/dt)]$ then the equation of motion is

$$m\frac{d^2x}{dt^2}+f\left(v_0-\frac{dx}{dt}\right)+\alpha x=0. \tag{20}$$

If we now make the assumption that $f[v_0-(dx/dt)]$ can be expressed as

$$C-k\frac{dx}{dt}+k'\left(\frac{dx}{dt}\right)^3,$$

then changing the variable so as to eliminate the constant term one obtains an equation of the form

$$m\frac{d^2x}{dt^2}-k\frac{dx}{dt}+\frac{1}{3}k'\left(\frac{dx}{dt}\right)^3+\alpha x=0. \tag{21}$$

Equation (21) can be put in a much simpler form by making the substitutions, $t_1=\omega t,\ \omega^2=(\alpha m),\quad x_1=\omega\sqrt{(k'/k)}\,.\,x,\quad \epsilon=k/\sqrt{(\alpha m)}=k\omega/\alpha.$

It then becomes

$$\frac{d^2x_1}{dt_1{}^2}-\epsilon\left\{\frac{dx_1}{dt_1}-\frac{1}{3}\left(\frac{dx_1}{dt_1}\right)^3\right\}+x_1=0. \tag{22}$$

We shall see that the character of the vibration is determined by a single parameter ϵ.

Although eqn. (22) can be obtained in a much more satisfactory manner, the tentatively intuitive derivation given above is to some extent typical of vibration problems of this kind and illustrates one of the difficulties, namely, that of obtaining an equation of motion which represents the system fairly accurately and which is at the same time not so cumbersome as to be unmanageable. A number of systems permitting self-sustained oscillations can be fairly accurately described by an equation of the above form. The best known electrical exampls is that of a triode oscillator in which a parallel *LC* circuit between grid and cathode is coupled to the anode through a mutual inductance. Perhaps the most famous mechanical example is provided by the action of a bow on a violin string. Here the friction between the slowly moving bow and the string causes the latter to vibrate due to variability of sliding friction between bow and string. Other examples such as the squeaking of chalk on a blackboard have been mentioned earlier.

Equation (22) cannot be solved directly and one must use analytical methods instead. One such method is the following. Dropping all subscripts we may write eqn. (22) in the form

$$\frac{dv}{dx}=\frac{\epsilon\{v-(v^3/3)\}-x}{v}=\frac{\epsilon F(v)-x}{v} \tag{23}$$

where v is the velocity dx/dt. (We may often be more interested in velocity than in displacement, particularly in electrical circuits since in this case displacement is analogous with charge and velocity with current, with which we are usually primarily concerned.) We first draw the curve $x=\epsilon F(v)$ as shown in fig. 7. A point $P(x, v)$ in the x, v plane specifies the velocity and position at any instant. We wish also to find the slope dv/dx at P. To do this a line from P is drawn parallel to the x axis until it cuts the curve at R. From R a perpendicular to the x axis is drawn intersecting it at S. Then we have $RS=v$, $PR=x-\epsilon F(v)$. A line aPb perpendicular to SP has slope $\frac{\epsilon F(v)-x}{v}$ at P and so satisfies eqn. (23). We now repeat the process for a point Q on the line aPb and by a series of such processes obtain a polygonal curve which can be made as good an approximation to a smooth curve as we wish by making the segments PQ shorter and shorter. In this way we obtain a complete set of curves showing how the velocity varies with the displacement. The time factor can be obtained from such a curve since $v=dx/dt$ and so $dt=dx/v$. Thus the x, v curve may be used to obtain either the displacement or the velocity as a function of time.

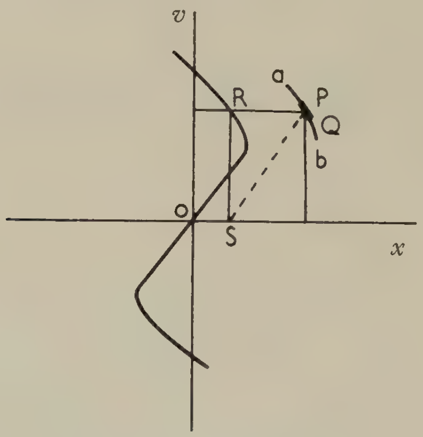


Fig. 7. Construction for obtaining v, x curves.

In the present example the spirals outwards from the origin or inwards from a point well removed from the origin, ultimately becoming asymptotic to a closed curve. Once the motion can be described by a closed curve we know that it is periodic. The character of the periodic motion is determined by the shape of the closed curve and this in turn is determined by the magnitude of ϵ. A typical v, x curve for $\epsilon \sim 2$ is shown in fig. 8. When ϵ is zero the characteristic is a straight line coincident with the v axis and the v, x curve is a circle of indeterminate radius. This corresponds to the case of free vibration without damping. When ϵ is finite but very small, the closed portion of the x, v curve approximates to a circle and the motion is therefore approximately simple harmonic. It is interesting to note in passing that this shows that a self sustained oscillation is never in the strictest sense, simple harmonic. A valve oscillator for example can never generate an absolutely pure sine wave; neither can we sing a pure tone. As ϵ is increased (fig. 9) the shape of the closed v, x curve departs more and more from circular form and this is reflected in an increased 'jerkiness' of the motion. When ϵ becomes very large the motion is no longer even approximately simple harmonic and the very jerky oscillations which occur for values of ϵ greater than about 10 are usually referred to as relaxation oscillations. The

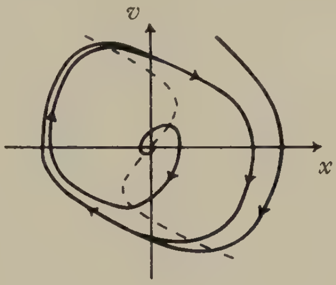


Fig. 8. Typical v, x curve for $\epsilon \sim 2$.

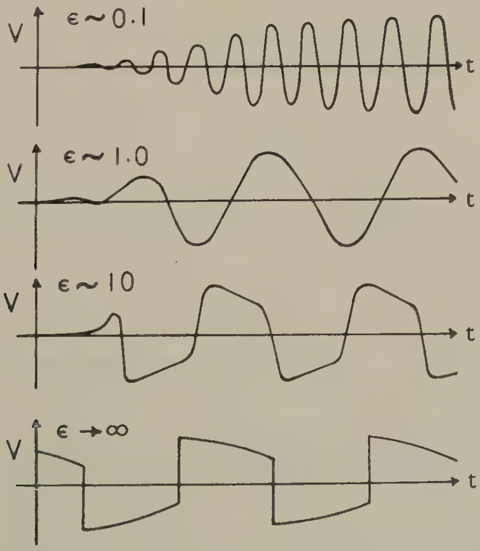


Fig. 9. The change in the character of the oscillations with increased non-linearity.

name arises from the fact that the oscillations exhibit two distinct phases; one in which energy is slowly stored up in the spring (or in the condenser in the case of an electrical circuit) and one in which the energy is discharged almost instantaneously once a certain critical value is reached.

There are good reasons for believing that self-sustained oscillations of the type considered above are of widespread occurrence. Van der Pol, who first obtained solutions of eqn. (22) in connection with the triode oscillator, showed that the heart beat is a relaxation oscillation. Another example from biology is that of the variation in the number of individuals in animal species which live together with one species feeding on the other. The problem is simplest when only two species are considered, for example rats and voles. If the voles existed alone they would multiply indefinitely. However the rats eat the voles, and as long as the supply of voles is adequate the rats multiply. Eventually the rats' demand for voles is sufficient to cause a rapid decrease in the vole population. The rats thus find themselves in a position in which their increasing number is accompanied by a rapid fall in the number of voles. The rats die off in large numbers due to lack of food and the vole population increases again. The number of rats and voles varies periodically and the variation shows distinct relaxation characteristics. A simpler example is that of a single herbivorous species. The animals multiply until they exhaust their own food supply, whereupon they either migrate and/or die off in large numbers. The great lemming migrations which occur every four years or so may be attributable to this cause. Large periodic variations in animal population are known to occur in many different species and it is highly probable that this is due to the influence of the population on its own food supply which causes self-sustained oscillations of the type considered above. Variation in human population, and the recurrence of epidemics and plagues may also be subject to the same characteristics.

In addition to biological systems there are a great many mechanical systems which can perform oscillations of this kind. The generation of oscillations by a moving air stream is closely connected with the formation of eddies which, under certain conditions can produce an effect similar to that which would be produced by negative damping. The most famous example of a mechanical system set in vibration by wind is that of the suspension bridge at Tacoma Narrows, U.S.A. This bridge was set into oscillations so violent that at times the transverse section of the roadway was ±45° to the horizontal with a vertical displacement of 28 ft. which eventually brought about its collapse in a wind of little more than 40 miles an hour. An interesting illustrated account of the failure of the Tacoma Bridge is given in *Engineering*, **150**, 481 (1940).

REFERENCES

FLETCHER, H., 1953, *Speech and Hearing in Communication* (New York: Van Nostrand).
HELMHOLTZ, H. VON, 1954, *On the Sensations of Tone* (New York: Dover Publications Inc.).
McLACHLAN, N. W., *Ordinary Non-linear Differential Equations* (O.U.P.).
STOKER, J. J., 1950, *Non-linear Vibrations* (New York: Interscience Publishers Inc.).
WOOD, A., 1940, *Acoustics* (London: Blackie).

THE AUTHOR is a Senior Lecturer in Physics at the University of Sheffield. His chief interests are in solid state physics, particularly magnetic properties.

The Post Office Research Station, Dollis Hill, London, N.W.2 Open days 22–23 September 1960

by V. T. SAUNDERS
Uppingham

When Oliver Cromwell's decision to place the Post Office under the control of Parliament was ratified by Charles II in 1660, the main business of the Post Office was communication—transmission of the written word by letter. Since these days the Post Office has become responsible for many other services, either on its own account, or as an agent for other Government Departments; communication, however, is still its main business, although many new forms of it have been added during the last hundred years.

The first new form was that of the transmission of the written word in coded form by the electric telegraph; this was followed by transmission of the spoken word by telephone. Transmission of pictorial information came next: first in the slow-speed form of picture telegraphy, then in the high-speed form of television. The latest form is data transmission, the data being mostly numerical, which is associated with the growth of electronic computers over the last decade or so; the administrative and technical arrangements for data transmission by the Post Office are still in the very early stages, but this service will undoubtedly grow.

The Open Days were mainly concerned with mechanical, electrical and electronic engineering—system work and equipment design—but amongst the exhibits there were examples of fundamental research and studies of particular problems. There is continuous study of speech but even so, it is said, it is still not possible to predict how a system whose treatment of regular sine waves over the entire audio and radio spectrum is known to extreme accuracy will strike two subscribers in conversation.

The metallurgical section furnished several examples of fundamental research. Nickel-iron alloys have been much improved by powder metallurgy and careful annealing, some success has been achieved with aluminium alloy and some metals have been rolled down to a thickness $\simeq 10^{-4}$ in.—this is of importance in reducing eddy current losses at very high frequencies.

Work on extremely clean surfaces such as can be retained for short times only in the best vacua is beginning as well as work on real surfaces.

Investigations of the wanted and unwanted physical effects in components and materials often lead to new effects which the members of the staff are encouraged to explore. At all stages problems in solid state physics and surface physics and chemistry are encountered; hence the need for a continuous supply of young physicists and chemists; there is an interesting and rewarding career here.

During a three-hour visit to an Exhibition having about a hundred exhibits the visitor can do no more than sample and report.

1. An electronic telephone exchange

This was a model of an electronic exchange to be put into service at Highgate Wood in 1962; it showed a large development project in which the Post Office enlisted the assistance of five outside firms. In the course of discussion it emerged that the power in the calling circuit of this exchange is not sufficient to ring a bell and therefore it will be necessary to devise a new method of calling the subscriber.

The very high operating speeds of electronic switches encourage new switching techniques, time-sharing the switches among a number of conversations, and requiring far fewer switches.

The switches must be capable of transmitting speech with a frequency range of rather less than 5 kc/s. This limited range makes it possible to sample the speech every 100 μs, to transmit the information through several switching stages and to reconstruct it without loss of intelligibility or quality. The sampling and transmission can be achieved in less than one microsecond. Thus, the switching stages need to be closed for only one out of every 100 μs and during the remaining time can be used for other speech signals. By this means the speech from 100 subscribers can be concentrated on to a single highway and be redistributed as required. This principle of cyclically allocating time intervals to a number of different sources is known as *time-division-multiplexing* (t.d.m.) and has been previously used in radio transmission.

The signal transmitted each 100 μs consists of a pulse of about $\frac{1}{2}$ μs duration whose amplitude is determined by the instantaneous value of the speech voltage at the moment of sampling. This system is known as *pulse amplitude modulation* (p.a.m.). The modulator consists of a simple diode gating circuit and the demodulation comprises a circuit for increasing the energy of the pulse by lengthening it in time and a filter for separating the audio frequency from the pulse repetition frequency (10 kc/s), its sidebands and harmonics.

In the electronic exchange a pulse train, corresponding to a time interval in the t.d.m. system, is allocated to a subscriber only when he indicates, by removal of his receiver, that a call is required. Depending on the calling rate of the subscribers it is possible to concentrate up to 1000 subscribers on to a single 100-channel t.d.m. highway. This requires a memory which will continue to supply a subscriber, for the duration of a call, with the pulse train that has been allocated. The memory consists of a group of 15 magnetostriction delay line stores with regenerative amplifiers. Each store is 100 μs long and can store 100 pulse trains. The stores are divided into three sets of five and when a pulse is inserted into two stores in each set the outputs are combined in a matrix so that only one subscriber receives the pulse train. There are $({}^5C_2)^3$ or 1000 ways of inserting the pulses. Where an exchange serves more subscribers than can be accommodated on a single highway, several highways are assembled and interconnection between the highways is done by additional switching stages also controlled by pulse stores.

In addition to the speech transmission, the exchange must receive the digital information from the subscriber's dial, and facilities such as ringing, metering, and routing of junction calls must be provided. These facilities involve only digital functions, and the equipment involves storage and circuits performing logical functions; these also use the t.d.m. principle. The outputs from the storage in this control section of the exchange can be less frequent than the

100 μs dictated by the speech, and more economical delay line stores with longer access times can be used. Some sections of the control equipment, e.g. the meter record, can operate with an access time in the millisecond range and a magnetic drum memory is employed.

The exchange thus comprises 100-channel t.d.m. highways and switches carrying pulses which are amplitude modulated by speech, the switches being controlled by logical circuitry including fast access storage. It also contains a digital section which processes the information received from the subscribers' switch-hooks and dials; this section involves memories operating with access times appropriate to their function.

2. Speech transmission using Pulse-Code Modulation

In this demonstration the p.a.m. of the electronic exchange was converted into *pulse-code modulation* (p.c.m.).

For the transmission of speech signals over the lines between exchanges p.a.m. would be unattractive. Relatively low-noise lines, having a linear phase response over a wide frequency band would be required in order to give low-noise speech channels with reasonable freedom from inter-channel crosstalk. Pulse-code modulation requires the transmission of binary pulses only, and offers important advantages.

The waveform to be transmitted is first sampled to produce p.a.m. pulses. Each p.a.m. pulse is then considered to have an amplitude equal to one of a finite number of amplitude steps, and a sequence of coded pulses is generated to identify each step. Thus if a straightforward binary code is used 2^n steps may be identified by n pulses. The signal obtained after decoding comprises a replica of the original waveform and a noise component due to quantizing.

As each p.a.m. pulse has been replaced by n binary pulses, the bandwidth required for their transmission is greater. The advantage offered by p.c.m. to offset this disadvantage, is that, as the p.c.m. pulses are binary, regeneration may be employed at each amplifier or ' repeater '. Provided that the repeaters are spaced at sufficiently frequent intervals, so that the presence or absence of a pulse may be determined unambiguously despite accompanying noise and crosstalk, the output of each repeater can be a perfect replica of the sent pulse train. The only noise accompanying the decoded signal is then that due to quantization; this may be set at any desired value by suitable choice of parameters for the terminal equipment.

Pulse-code modulation is an attractive technique for use when a transmission medium offers large bandwidth at relatively low cost, but is subject to serious noise and interference. Such a situation arises when attempts are made to increase the carrying capacity of inter-exchange multi-pair cables, at present inductance-loaded and carrying audio-frequency signals. If the loading coils are replaced by repeaters a transmission bandwidth of about 1 Mc/s can be obtained; but crosstalk between pairs severely limits the use of these cables for multi-channel transmission by analogue methods. Nevertheless they may be acceptable for digital signals. In this application, which seems likely to provide the first use of p.c.m. in the public telephone network, the very wide range of signal levels ($>$ 30 dB) incoming to the p.c.m. system under different conditions is a complicating factor. It is known that the use of a non-uniform distribution of coder levels, favouring the low-level signals, is helpful in this connexion, but much remains to be done to determine the appropriate numbers of levels, and their distribution.

The Post Office has a research programme which includes the study of circuit techniques for the coding and decoding of speech signals, and subjective testing to establish the desirable parameters of p.c.m. systems. An experimental p.c.m. terminal equipment demonstrated on Open Day, is used in connection with these experiments. It offers 12 telephony channels, each sampled at 8 kc/s, and the number of digits available to identify each speech sample may be varied up to eight. By making circuit adjustments, different distributions of coder levels may be set up for experimental purposes.

The experience gained during this work will also form a background for the exploitation of p.c.m. in other applications, such as arise, for example, in connection with long-distance circular waveguides.

The writer received speech based on an 8-digit and on a 3-digit code level; the former was full-bodied and pleasing and the latter thin and meagre—the difference between a vintage wine and distilled water!

3. Audio spectrum analysis in ‘ real time ’

Here was speech made visible, in detail, on a c.r.o.—a new alphabet-cum-spelling all in one. The demonstrator who spoke and hissed and hummed and whistled into the microphone said that this picture of sound can indeed be read and that one of the future developments of this may enable the totally deaf to ‘ read-hear ’ current conversation.

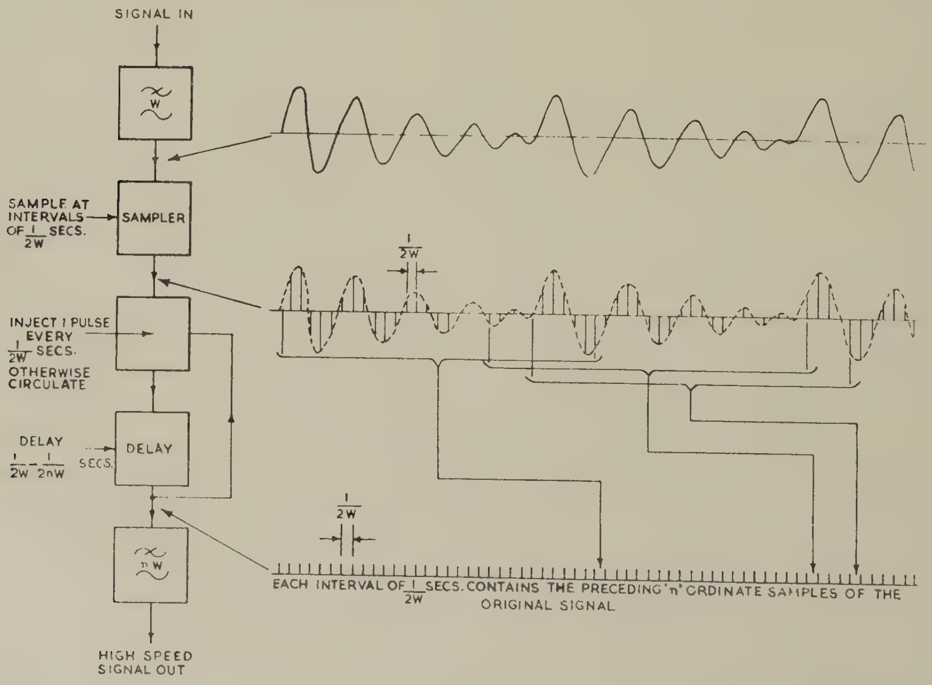

SIGNAL IN
W
SAMPLE AT INTERVALS OF $\frac{1}{2W}$ SECS.
SAMPLER
$\frac{1}{2W}$
INJECT 1 PULSE EVERY $\frac{1}{2W}$ SECS. OTHERWISE CIRCULATE
DELAY $\frac{1}{2W}-\frac{1}{2nW}$ SECS.
DELAY
$\frac{1}{2W}$
nW
EACH INTERVAL OF $\frac{1}{2W}$ SECS. CONTAINS THE PRECEDING 'n' ORDINATE SAMPLES OF THE ORIGINAL SIGNAL
HIGH SPEED SIGNAL OUT

Fig. 1. High-speed repetition of the immediate past of a signal by circulating memory store.

There are many calls for spectrographic analysis in ‘ real time ’, that is to say for a method of deriving a spectral description which unfolds as the signal proceeds. A bank of electrical band-pass filters and rectifiers can provide this facility but the method becomes impractical when large numbers of filters are needed. For example, at least 160 filters are needed to display 50 c/s analysis of a 4 kc/s wide band.

When operation in ‘ real time ’ is not required the analysis can be realized quite economically and flexibly. The signal is recorded and subsequently analysed by repeated playback through a single analysis filter which is swept slowly (to avoid errors due to filter ‘memory’) across the signal band. This is the method used in the sound spectrograph. The time required for analysis (typically several hundred times that needed for recording) can be reduced by increasing the rate of signal repetition but ‘ real time ’ working can never be achieved unless recording and high speed repetition proceed simultaneously.

Continuous high speed repetition of the immediate past of a signal, of bandwidth 0–W c/s, can be realized, in principle, by the method shown in fig. 1. A circulating delay line memory is arranged to accept samples of the signal taken at $1/2W$ sec intervals and store them at $1/2nW$ sec spacing. During each $1/2W$ sec interval the entire contents, namely the last n signal samples, are available for analysis. Each sample is presented during n successive circulations and so up to n analyses can be performed by a single filter in the available time.

Transient effects in the analysis filter caused by the $1/2W$ spaced discontinuities in this high speed signal must be removed from the final display. This is possible provided that the majority of the impulse response (that is to say the region in which the response exceeds $1/10$ of its maximum value) is contained within $1/2W$ sec.

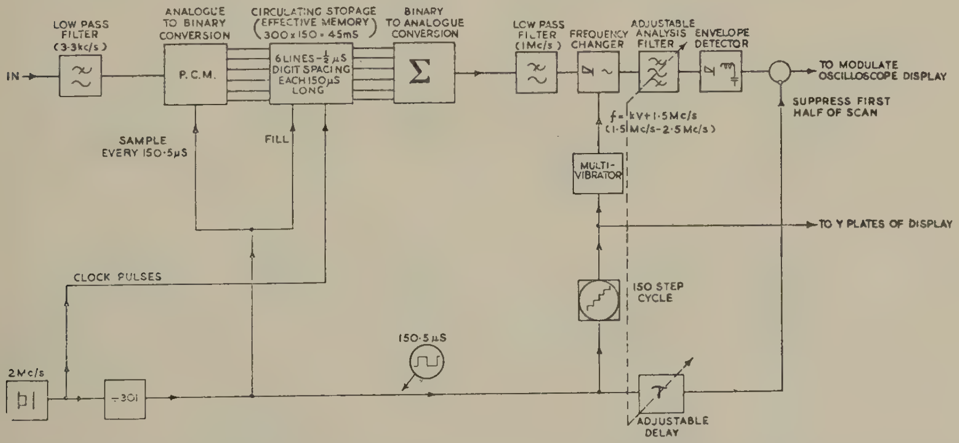

LOW PASS FILTER (3·3kc/s)
ANALOGUE TO BINARY CONVERSION
CIRCULATING STORAGE (EFFECTIVE MEMORY 300 x 150 = 45mS)
BINARY TO ANALOGUE CONVERSION
LOW PASS FILTER (1Mc/s)
FREQUENCY CHANGER
ADJUSTABLE ANALYSIS FILTER
ENVELOPE DETECTOR
IN
P.C.M.
6 LINES-½ µS DIGIT SPACING EACH 150µS LONG
Σ
TO MODULATE OSCILLOSCOPE DISPLAY
SUPPRESS FIRST HALF OF SCAN
f = kV+1·5Mc/s (1·5 Mc/s-2·5 Mc/s)
SAMPLE EVERY 150·5µS
FILL
MULTI-VIBRATOR
TO Y PLATES OF DISPLAY
CLOCK PULSES
150 STEP CYCLE
150·5 µS
2Mc/s
÷301
ADJUSTABLE DELAY

Fig. 2. A short-term spectrum analyser capable of processing a 3·3 kc/s wide signal band in ‘ real time ’ at analysis bandwidths (half-power) down to 47 c/s.

For an idealized filter of gaussian form if f is the interval between ± 3 dB points on the frequency response characteristic and t is the interval between ± 20 dB points on the corresponding impulse response, then $ft \simeq 1{\cdot}14$.

It can be shown that, for the above method, the spacing (τ) of the samples within the memory required for spectrum analysis of a signal of bandwidth W at analysis bandwidths (gaussian) down to f is:

$\tau \simeq f/4{\cdot}5W^2$ for a display of spectral cross-sections

and $\tau \simeq f/9W^2$ for a normal three-dimensional spectrogram.

The circulating memory can be a magnetic or dielectric drum store if a low sampling rate, say less than 500 / sec, is acceptable, but for wide band signals such as speech no suitable method for circulating the information in analogue form has yet been realized. At present, the only effective solution is to use a binary digital storage system.

Continuous spectrum analysis of a 3·3 kc/s wide band at analysis bandwidths (gaussian) down to 47 c/s has been achieved by the method outlined in fig. 2. A representative set of spectrograms produced in this manner is shown in fig. 3.

The instrument exhibited displayed a continuous spectrogram by means of a long-persistence oscilloscope. Visitors were thus able to see how, for instance, the vocal resonances shape the energy–frequency spectrum of speech.

4. STUDIES OF SILICON AS A TRANSISTOR MATERIAL

The investigations into germanium-alloy p–n–p transistors have been extended to silicon-alloy p–n–p types; the visitor was shown surfaces and junctions under

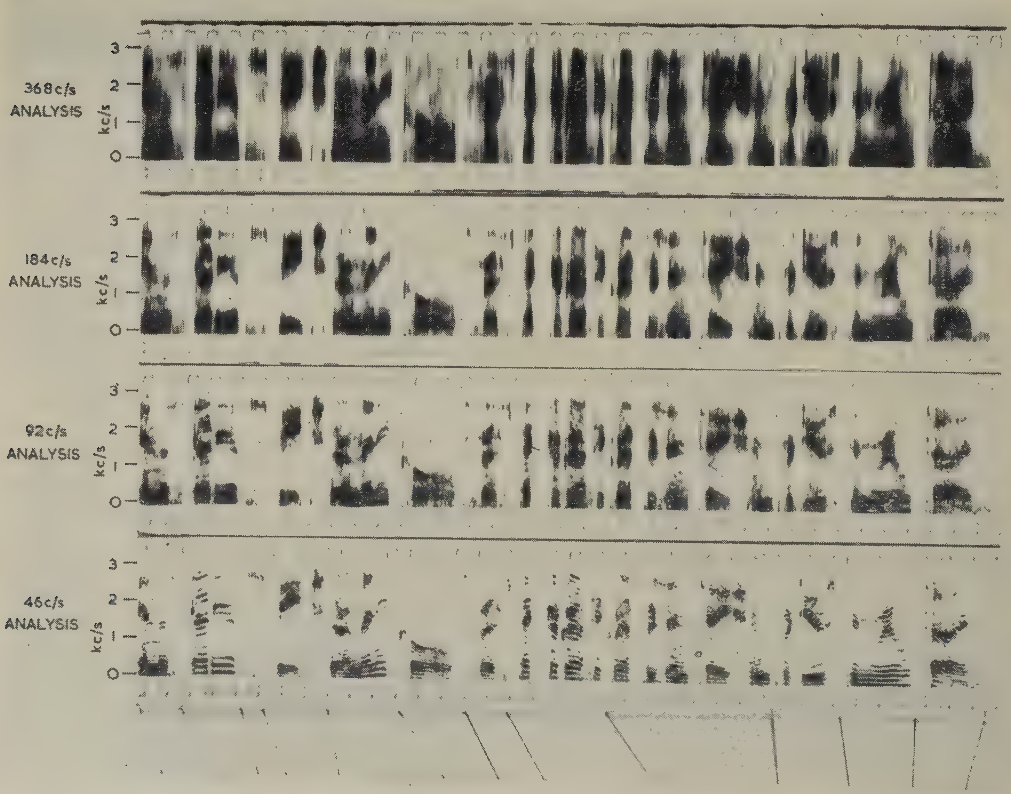


Fig. 3. Continuous spectrograms, produced by the analyser shown fig. 2, of the same spoken sentence at different bandwidth analysis.

the microscope and there was mention of a *p*-type silicon having 5×10^{15} atoms of boron per cm^3.

Large electronic equipments are practical only if the reliability of many components can be ensured. The failure mechanisms of transistors have still to be fully separated and described in physical, chemical and metallurgical terms. Many properties of the semiconductors used, and the processes whereby *p–n* junctions and other inhomogeneities are deliberately introduced in the making of a transistor, must be examined to see whether, when the device is used, atoms in the structure proper or in the surrounding atmopshere can move in such a way as to change the electrical properties.

Thus a *p–n* junction made by diffusing phosphorus to a depth of about 10 microns into *p*-type silicon was revealed by bevelling at a very small angle ($\simeq 2°$) and staining the *n*-region in order to examine the fine structure, if any, of the contour of the junction and its dependence on the process used to render the surface of the starting material optically flat; in fact no irregularities could be seen when the polishing involved little mechanical damage. These polished surfaces are often oxidized, both to produce a selective mask against phosphorus diffusion and as a protective surface on a completed transistor. Electron micrographs ($\times 40\,000$) of replicas of some thermally grown oxide layers (thickness $\simeq 1\,\mu$) showed uneven oxidation resulting from minute scratch marks left by one polishing process, but others, of surfaces prepared with greater care, showed none. Although the germanium and silicon used to make transistors and power rectifiers are the purest and most perfect crystalline materials known to man, they can still contain dislocations of importance to the initial performances and possibly to the stability of devices made from them. Of

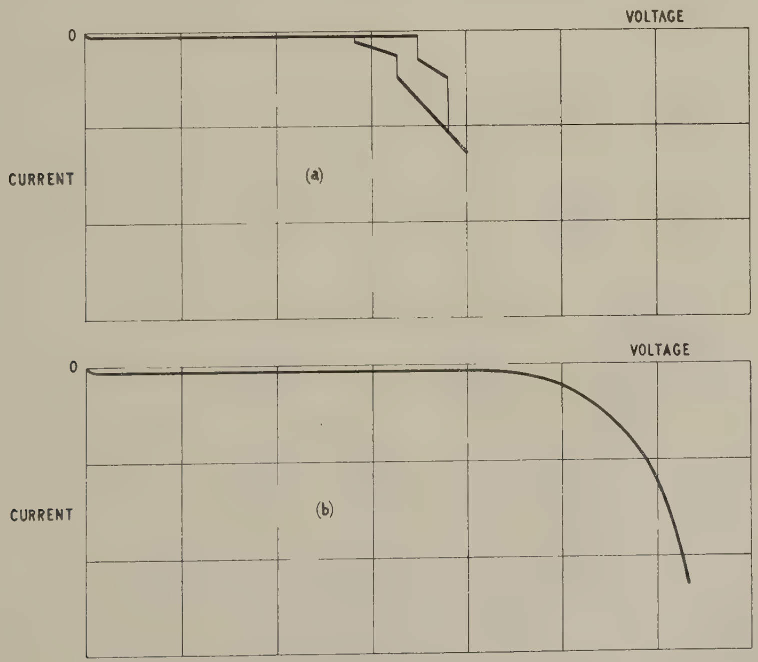


Fig. 4. (*a*) Reverse–current/voltage relationship of diode made of silicon containing dislocations.
(*b*) Reverse–current/voltage relationship of diode made of dislocation-free silicon.

particular interest is the electrical breakdown at reverse-biased junctions which results when the applied voltage is so increased that the electrical field can now give the few charge carriers (responsible for the small reverse current at low biases) sufficient energy in a few mean free paths as to cause ionization of atoms by collision, thereby increasing the number of carriers and ultimately producing a rapid rise in current by ‘ avalanching ’. The process is usually accompanied by light emission from discrete spots (about 100 A in diameter and called microplasmas) identified as dislocations running through the junction. Accompanying the appearance of each spot of light is a step of about 100 μA in the V/I relationship of the reverse biased diode, fig. 4 (*a*). Junctions made on dislocation-free material by a process which does not introduce strain, show no steps in the breakdown, fig. 4 (*b*), and no bright spots. The presence of the discrete regions carrying the avalanche effect appears to have no short term harmful effects, but the long term stability could still be marginally affected even though, in use, the voltages applied are much less than those needed to cause avalanche effect.

5. Sorter resorted

The work already described is directed to improving existing methods and more efficient machines. But ‘ postal mechanization ’ is something comparatively new and its object is the replacement of men by machines; we, and most other industrial countries, may soon be short of men and women who are prepared only to sort letters—hence the urgency of this work.

Exhibit No. 26. *Organic Phosphors.* Demonstrated selectively excitable organic phosphors for coding stamps and letters to be handled by automatic sorting devices. *No.* 33. *Letter-mail Address-encoding Operator's Desk.* The operator sits at a desk with a key-combination, i.e. a typewriter pattern keyboard, and each letter is brought up before him on a belt; he reads and ‘ types ’ the first three and the last two letters of the destination town—for BIRMINGHAM he types BIRAM; by electronic translation dots representing in code form the destination of the letter are printed in phosphorescent material on the envelope which is then transported to exhibit *No.* 34. *Automatic Letter Sorter.* Here the phosphorescent address code, which is virtually invisible in normal light, is activated by ultra-violet light and read photoelectrically; finally the letter is carried by a belt over the top of a long row of place-named boxes each with a top flap door which opens when it receives instructions from the code reading unit.

Another exhibit which, by its startling name, called for attention, was *No.* 88. *MASSACRE—M*achine for *A*utomatic *S*urface *S*ampling and *A*utomatic *C*ontact *R*esistance *E*valuation. It is used in studying the effect of environment on the mechanical and electrical properties of surface films on contact materials. A plate two inches square can be tested at a thousand different points, each observation occupying only four seconds.

The writer has great pleasure in recording his thanks first to Dr. J. R. Tillman who patiently gave three whole hours to just one visitor and secondly to all those members of the staff of the Establishment who have supplied the descriptive notes on which this article is based.

Advancing Fronts in Meteorology

by SIR GRAHAM SUTTON, C.B.E.

Director-General of the Meteorological Office

1. Introduction

Meteorology is ' the science of the atmosphere '. Because its study embraces almost the whole of classical physics, and especially the mechanics of fluids, heat conduction, radiation, optics and electricity, as well as some aspects of physical chemistry, it has tended in recent years increasingly to separate into specialized branches. The broadest and oldest division is into ' synoptic ' and ' physical ' meteorology. Synoptic meteorology (which may be understood here to cover also dynamic meteorology) is difficult to define in precise terms, but it is generally understood to imply the study of atmospheric dynamical systems of large horizontal length-scale considered in a time-sequence. Physical meteorology is concerned with the details of the processes that produce the characteristic features of the atmosphere. There are now many subdivisions, and meteorologists may specialize, for example, in cloud physics, with special emphasis on the problems of condensation, in micrometeorology, which is largely preoccupied with small-scale transfer processes in the lower layers of the air, or in atmospheric chemistry, which deals with the chemical composition of air masses and aerosols, and their geographical distribution.

The forecasting of weather is a professional activity which is related to the science of meteorology much in the way that the art of healing depends upon the medical sciences. The analogy between meteorology and medicine is close in many respects, for just as all qualified doctors spend part of their early years in clinical work, so the great majority of meteorologists experience the discipline of the forecasting bench. It is most unlikely that a new technique in weather forecasting would be introduced by a national service without prior examination and development by the scientific meteorologists, just as a new treatment in medicine would be applied only after approval by medical research workers. The ' routine forecaster ' of the national weather services is the counterpart of the ' general practitioner ' in medicine, and in the mind of the public his pronouncements are regarded as the main justification for the cost of the service. But this is a partial view, for the less widely known advisory services of meteorology, which often have little connexion with forecasting, contribute to no small extent to the economy of a country.

To see the science of meteorology in proper perspective in the general field of physics, and to appreciate the nature of the very substantial advances that are now being made, requires a knowledge of the characteristics of the medium in which the meteorologist works—the atmosphere of the earth—and also something of the historical development of the subject. We begin with a brief account of those features of the atmosphere that are the main concern of the meteorologist with an indication of the facts that call for explanation.

2. The atmosphere

During the final stages of its evolution, the earth became a solid body with a surface which is about three-quarters water, and a thin gaseous envelope, composed mainly of nitrogen, oxygen and water-vapour. This skin, the atmosphere, accompanies the planet in its orbit around the sun and, on the whole, rotates with it.

For the meteorologist, the most important constituent of the atmosphere is water-vapour, despite the fact that it accounts for only a small fraction of the total mass of terrestrial air. The other gases of the atmosphere, which make up by far the greater part of its mass, are inert in the meteorological sense. The overriding importance of water in meteorology arises from a number of causes, such as its very large specific and latent heats (especially the latent heat of vaporization), its ability to absorb and re-emit infra-red radiation freely and, especially, in the circumstances of its terrestrial location, its liability to frequent changes of state. It is commonplace that water in a shallow pool may freeze in the night and evaporate completely the next day. No other naturally abundant substance behaves like this or has similar properties. Water in the atmosphere stores and releases very large amounts of energy in the form of heat, and the characteristic phenomena of weather (which as far as we know exist only on the planet earth) arise, in the main, from the relative abundance of water in the lower levels of the atmosphere. The meteorologist is often justified in regarding air simply as diluted water-vapour.

The atmosphere is irradiated by the sun, the greater part of energy so received being in the form of electromagnetic waves of length between $0{\cdot}29\mu$ and 2μ. This band includes the visible part of the spectrum ($0{\cdot}4\mu$ to $0{\cdot}76\mu$), but about half the energy is in the near infra-red. Solar radiation, emerging from a body at about 6000 °K, is usually referred to in meteorology as ‘ short-wave radiation ’, in contrast to the radiation emitted by the earth and the atmosphere which, originating in a body at about 300 °K, is entirely in the far infra-red, with wavelengths between 3μ and 80μ. Terrestrial and solar radiation do not overlap in wavelength to any appreciable extent, and radiation originating in terrestrial sources is invariably called ‘ long-wave ’ in meterological literature.

The heat output of the sun, judged by the measurements of the past century, is remarkably steady, with an intensity (the *solar constant*) assessed at very nearly $2\ \mathrm{g.\,cal\,.\,cm^{-2}\,.\,min^{-1}}$ at the fringe of the atmosphere when the earth is at its mean distance from the sun. About 85 per cent of this energy can pass through a clear atmosphere to reach the surface of the earth, but on the average about 40 per cent is reflected back to space by clouds, seas and lakes, vegetation, snow and soil. This fraction is called the *albedo* of the earth and its atmosphere.

Solar radiation is the only source of energy for the atmosphere, contributions from the hot interior of the earth and from other heavenly bodies being insignificant in comparison with that from the sun. The evidence of history and the analysis of meteorological observations for the last century or so show that, on the whole, the earth is neither gaining nor losing heat at an appreciable rate. It follows that over a long period of time the entire input of energy from the sun must be returned to space. This is the *heat balance* between the atmosphere and the sun.

The details of the balance sheet are complicated and subject to many uncertainties. A simple calculation shows that if the earth, radiating to space as a black body, were in exact balance with 60 per cent of the solar output (taking the albedo to be 0·4), its average surface temperature would be about -27 °c. The observed mean temperature of the surface of the earth is close to 15 °c, so that much of the long-wave radiation emitted by the earth is absorbed by the atmosphere and re-emitted to keep the surface and the lower air layers warm. The absorption is mainly by water-vapour and, to a lesser extent, by carbon dioxide. This ' greenhouse effect ' underlines the important part played by atmospheric water in the heat economy of the atmosphere.

A detailed examination of the heat balance shows that from the equator to about latitude 40° incoming radiation exceeds that lost to space, and that from altitude 40° to the poles, the reverse is true. Without a compensating mechanism the tropics would grow steadily warmer and the polar areas more frigid. The compensation is achieved by the internal motions of the atmosphere (winds) and to a lesser extent by ocean currents. The quasi-permanent pattern of wind which is revealed by averaging observed winds over long periods is called the *general circulation of the atmosphere.* Its study is a major interest in present day meteorology.

The strong internal motions of the atmosphere, together with the frequent changes of phase of the water content of the air, give rise to the characteristic features of *weather.* (In meteorology, weather is defined as the state of the atmosphere at any time—especially, its temperature, humidity, winds, cloudiness and precipitation.) It is a familiar fact that weather changes rapidly with time. The average weather of a region is its *climate.* A study of the motion of the atmosphere by means of averages taken over varying periods of time is equivalent to a filtering process in the statistical sense. At heights of the order of a few metres above ground the wind is usually highly turbulent, with oscillations of periods varying from fractions of a second to many minutes. The process of taking averages over periods of the order of an hour removes the more rapid oscillations and leaves a record with large slow surges. Usually, these variations in the mean-wind speed are indicative of the passage of well-defined dynamical systems, which are known as depressions (cyclones or lows) and anticyclones (highs). The systems are easily recognized on charts of atmospheric pressure, and a succession of such charts shows that they travel, usually from west to east, at speeds up to a few tens of kilometres an hour. The art of weather forecasting, as practised at present, depends largely upon the persistence of such systems as distinct features of the weather chart for several days at least, and the fact that certain characteristic types of weather are associated with them.

In the extra-tropical latitudes there is a well-defined relation between the pressure and motion fields. The driving force of the wind is provided by horizontal pressure gradients, but in large-scale atmospheric dynamic systems the motion of the air is not simply ' down the gradient ', i.e. from high to low pressure. The rotation of the earth provides an apparent steering force, depending upon the so-called Coriolis acceleration, which deflects the moving air to the right in the northern hemisphere and to the left in the southern hemisphere. At heights greater than about 500 metres, where the effects of surface friction are negligible, the pressure gradient and the Coriolis forces come almost into balance to produce a close approximation to the theoretical *geostrophic wind*, which blows along the isobars at a speed proportional to the magnitude of the pressure gradient and inversely proportional to the sine of the latitude. Large-scale atmospheric systems are very nearly in geostrophic balance and the isobars approximate closely to streamlines, so that, in practice, it is often sufficient to plot only the pressure field, leaving the motion field implied.

The range of atmospheric motion systems is very extensive, from great planetary waves and continental anticyclones with horizontal length-scales up to tens of thousands of kilometres, to the familiar depressions of the mid-latitudes (horizontal dimensions in thousands of kilometres), tropical hurricanes (diameter hundreds of kilometres) down to the minute convectional and frictional eddies, of size varying from metres to millimetres, that bring about transfer processes in the lower atmosphere. The task of the research meteorologist is to bring these systems and their associated physical features into the ambit of consistent physical theories.

The field of meteorological investigation is large, and it is possible here to select only certain topics, which seem to the writer to show signs of rapid development at the present time, for discussion. But the whole front of meteorology is on the move, and it is difficult to separate one advance from another both as regards significance and intellectual appeal.

We begin with the most accessible part of the atmosphere, in which life is most abundant and the physical processes most complicated.

3. THE LOWER ATMOSPHERE

Micrometeorology, which is mainly concerned with the study of the atmosphere close to the ground, is largely a product of the present century. It is obvious to anyone who has looked at an anemometer record, or even watched a wind vane on a church steeple on a day in summer, that on most occasions the motion of the air is highly turbulent, that is, made up of a mean velocity with large irregular variations both in speed and direction. Such oscillations play an important part in the meteorology of the lower atmosphere, for they are mainly responsible for the diffusion of momentum, mass and energy from one level to another. In a fluid at rest or in laminar motion, diffusion is entirely caused by molecular agitation, and in consequence is a slow process. In a turbulent fluid diffusion is attributed primarily to the ‘ eddies ’ in the flow, which in this respect act like molecules, but they bring about greatly increased rates of transfer, and the diffusing action of turbulence far outweighs that of molecular agitation.

In the motion of a homogeneous fluid through a pipe or in shallow layers over a solid surface, the onset and many of the typical features of the turbulence depend solely upon the value of the Reynolds number. In the lower atmosphere the situation is complicated by the presence of gravitational and compressibility effects arising from diurnal changes in the vertical gradient of density. In daytime, with a clear sky, the ground is strongly heated by the solar beam, with the result that air temperature falls rapidly with height in the first few metres above the surface. Such a marked gradient of temperature (called a *lapse-rate* in meteorology) indicates strong static instability, and the convection currents so formed maintain the wind in a highly turbulent state. In this condition there is a rapid exchange of momentum between the upper and lower layers of the air, with the faster moving eddies carried upwards and the slower air masses downwards. The same process also causes greatly enhanced conduction of heat and diffusion of suspended matter, such as water-vapour, dust and smoke. On a clear night the reverse condition holds. The ground now loses heat rapidly by radiation to space, so that very soon cold air lies below relatively warm air and the sign of the temperature gradient is reversed, a condition known to meteorologists as an *inversion.* In these circumstances the lower layers are statically stable and turbulence is suppressed. In the slow near-laminar motion of a clear night, diffusion of all three entities is at a low level. In particular, because of the reduced exchange of momentum, wind speed increases more rapidly with height in the inversion period than in the lapse period.

Atmospheric turbulence thus differs from that found in wind-tunnels and studied in aerodynamic laboratories in that, in the meteorological problem, density gradient exercises the controlling influence. The primary aim of many past and current micrometeorological studies has been to produce a hydrodynamical theory which relates diffusion in the air layers near the ground to easily measured features of the atmosphere, such as the speed and direction of the mean wind together with some observable parameters which are related to the intensity of the turbulence.

The earliest attempts to formulate such a theory were based upon the analogy between the turbulent and molecular exchange processes. It was assumed that a turbulent wind possessed a virtual viscosity, conductivity and diffusivity specified by a constant coefficient K (called the eddy viscosity, conductivity, diffusivity) analogous to the kinematic viscosity, thermometric conductivity and diffusivity of the kinetic theory of gases but, naturally, of much greater magnitude.† In the initial stages of the investigations this assumption appeared to be well founded, for reasonably consistent values of K (of the order of 10^3 to 10^4 cm^2 sec^{-1}, compared with values of the order of 10^{-1} cm^2 sec^{-1} for the molecular constants) were deduced from phenomena as dissimilar as the formation of fog in the North Atlantic, the variation of the diurnal wave of air temperature with height and the approach of the surface wind with increasing height to the frictionless geostrophic wind. Later, and more detailed studies, however, showed that no such universal constant exists, and a wide range of K-values, of magnitude from 1 to 10^{12} cm^2 sec^{-1}, were found with phenomena of different horizontal length-scale. It became abundantly clear that eddy and molecular transfer processes differ fundamentally, and not merely in magnitude (1).

The more recent attempts to solve the problem are founded, for the most part, upon the statistical theory of turbulence, initiated by Sir Geoffrey Taylor in 1921 (2) and greatly extended by him in 1935, and later by A. N. Kolmogoroff (3). The starting point of Taylor's work is a remarkable but essentially simple relation between the ' wandering ' of a particle in a fluid and the ' Lagrangian ' correlation between its motion at one instant and at some time later. From this beginning it has proved possible, not only to gain much insight into the structure of a turbulent fluid but also (with some plausible empirical assumptions concerning the functional form of the correlation coefficient) to deduce expressions which relate the spread of gas or suspended particles to the wind speed and its rate of change with height. The vertical gradient of wind speed near the ground, for reasons given above, is a reliable indicator of the intensity of the turbulent mixing, and there is every reason to believe that momentum and mass are exchanged at much the same rate in a turbulent fluid.

---

† Some writers use the *austausch* coefficient A instead of K. The relation is $A = K\rho$ where ρ is the average density of the air.

The statistical theory of turbulence, in its latest form, is too elaborate to be discussed here in any detail, and only a brief summary of the main results can be given. There now exist semi-empirical expressions which predict, with tolerable accuracy, the concentrations of matter found in known meteorological conditions from given steady sources (such as smoke generators at ground level, or factory chimneys) up to several kilometres from the point of origin. The formulae are most reliable in conditions of small vertical gradient of temperature. The problem of evaporation from water surfaces limited in extent downwind but long across wind has also been solved by the same approach.

In recent years, much attention has been given to the dynamics of natural convection. The problem of the rate of spread of a plume of warm air rising from point and line sources of heat in a calm atmosphere has been solved to a useful degree of approximation. The more difficult case of a hot plume which is deflected towards the horizontal by the natural wind (as in the guttering of a candle flame in a draught) has also received much attention, and although it cannot yet be claimed that a satisfactory solution has been derived, considerable strides have been made towards the final solution. The transfer of heat upwards on a summer day is a mixture of natural and forced convection and it is now possible to give a reasonably satisfactory explanation of the change in amplitude and phase of the diurnal wave of temperature with increasing height.

These problems are not only fundamental in the meteorology of the lower atmosphere, but have applications in wider fields. The control of atmospheric pollution is now recognized as urgent, but little can be done in promoting legislation until there is a sound scientific basis for assessment and planning. The study of the ' bent-over plume ', referred to above, has found considerable applications in the design of factories and industrial plant. Studies of evaporation and of the transfer of heat to and from the ground are clearly of importance in hydrology and agriculture, and the list might be extended almost indefinitely.

The advance of micrometeorology in the first half of the present century has been remarkable. Before 1915, when Sir Geoffrey Taylor published his great memoir on eddy motion in the atmosphere (4), very little attention had been given to the special problems of the turbulent lower atmosphere. In 1925 Wilhelm Schmidt published his well-known monograph *Der Massenaustausch in freier Luft*, which contained much valuable data but, on the theoretical side, was entirely devoted to the ' constant-K ' theory. The succeeding years saw the gradual abandonment of this hypothesis and its replacement by more sophisticated but still semi-empirical concepts derived, for the most part, from aerodynamic research, such as the ' mixing-length ' theory of Prandtl. At the same time considerable progress was made in the systematic examination and measurement of the properties of the lower atmosphere, especially by Sir Nelson Johnson and those who followed him at Porton. At present the lead seems to lie with the mathematicians. Most of the crude empiricism of the early years has disappeared, but some of the latest and most interesting work is, as yet, too abstract to allow of immediate application to the problems of the professional meteorologist (5).

The position which has now been reached in micro-meteorology can be summed up as follows: the limited technical problem, the prediction of diffusion by means of measurable entities, has been solved to the extent that techniques are now available to the working meteorologist which allow him to carry out the necessary calculations with a fair degree of confidence. The experimental problem, the exploration of the relations between the main features of the turbulent lower atmosphere, has attracted many workers and has resulted in some novel and ingenious instruments. With their aid, a great deal has been learned about matters such as the amplitude and frequency distributions of the oscillations. The most important and difficult problem of all, that of the construction of a rational theory of atmospheric turbulence, is part of the central theme of modern fluid dynamics research.

Micrometeorology is the branch of the science of the atmosphere which lies closest to experimental and theoretical classical physics. Although control of the environment and the isolation of selected processes, the essential features of all laboratory investigations, are absent, it is possible to make measurements of high accuracy and to test hypotheses in a way that is impossible in large-scale meteorology. A clear-cut result is more easily obtained in the problems of the accessible lower atmosphere than elsewhere and this, no doubt, accounts for much of the attraction of this rapidly growing branch of atmospheric physics.

4. SYNOPTIC AND GLOBAL-SCALE PROBLEMS

The atmosphere, a huge heat engine with the sun as the source of energy and water-vapour as the working substance, exhibits internal motions covering a wide range of horizontal length-scales. Much of the work of meteorologists in the present century has been concerned with the anatomy of the mid-latitude weather systems, with horizontal dimensions of the order of a thousand kilometres. The outstanding feature of meteorology in the period between the two world wars was without doubt the development of air-mass analysis and the wave theory of cyclones, on lines initiated by the Bergen school of Wilhelm Bjerknes. These researches established the main features of the typical weather-producing disturbances of the temperate zones and weather forecasting, as practised by the national services, depends essentially upon the recognition of air-masses and the movements of their boundaries, the familiar *fronts*. A great deal is now known concerning the dynamical and physical processes which give rise to the weather characteristic of depressions and their frontal zones (6).

Even so, progress in the forecasting of weather has been slow, for the problems are extremely complex and reliable observations have been none too plentiful. It is perhaps not generally realised by physicists that the accurate measurement of meteorological entities is often an extremely difficult matter. It is still not possible, for example, to measure the humidity of the air to anything like the degree of precision required by the physical meteorologist, and all instrumental problems become exceptionally difficult in the extremely dry, cold attenuated higher atmosphere. In these circumstances the advances that have been made must be regarded as notable achievements.

The better understanding of atmospheric processes that has come in recent years has also made clear some of the fundamental limitations of prediction in meteorology. It is now recognized that weather forecasts can never approach the precision and accuracy of those of spherical astronomy. An almanac of the atmosphere, a volume in which it would be possible to look up the weather of a locality some months ahead, is never likely to be produced. The nature of the motions of the atmosphere seems to preclude this possibility, and a weather forecast will never be more than a statement of probabilities.

The routine forecasting of weather, as carried out by the national services today, is a highly subjective process in which the experience and physical insight of the forecaster matter a great deal. In the preparation of a forecast there are many factors to be considered, of which relatively few can be specified mathematically. The traditional method of displaying the pressure pattern by the construction of isobars on level (constant-height) surfaces is now used professionally only for the sea-level map. Upper-air charts invariably show *contours*, or lines passing through points at which an isobaric (constant-pressure) surface has the same height above sea-level. For convenience, the scale of height used is the geodynamic metre, which differs only slightly from the ordinary metre. In addition, the height intervals between two isobaric surfaces, called *thickness*, are also plotted. Isopleths of thickness indicate the variation of the mean temperature of the stratum concerned.

The main mathematical aids are:

(1) the geostrophic-wind equations, $$u_g = -\frac{1}{\lambda}\frac{\partial h}{\partial y}\;;\quad v_g = \frac{1}{\lambda}\frac{\partial h}{\partial x} \tag{1}$$

where u_g and v_g are the components of the geostrophic wind along axes fixed in the earth, with OX pointing to the north and OY to the east, h is the contour-height of the selected isobaric surface (h, of course, is a function of x and y) and λ is the Coriolis parameter $2\omega \sin \phi$, where ω is the rate of rotation of the earth and ϕ is latitude;

(2) the thermal-wind equations, $$u_T = -\frac{1}{\lambda}\frac{\partial h'}{\partial y}\;;\quad v_T = \frac{1}{\lambda}\frac{\partial h'}{\partial x} \tag{2}$$

where h' is the thickness; and

(3) the hydrostatic equation, $$\frac{\partial p}{\partial z} = -g\rho \tag{3}$$

where z is height, p is pressure and ρ is density. This relation is sufficiently accurate (because of the small vertical accelerations) to be used in the calculation of heights from the radio-sonde observations. In addition there are certain theoretical results which act as guides. The most useful of these depend upon Sutcliffe's relations between the vorticity of the pressure and thermal motion fields and the development of the weather systems (7).

Routine forecasting by conventional methods proceeds by well-defined stages, beginning with the analysis of the initial situation from the observations, followed by the prognosis, which culminates in the freehand drawing of charts of the pressure and temperature fields as they are expected to be 12 or 24 hours later (a pressure-distribution prepared in this way is called a *prebaratic*†) and finally, a deduction of the weather (temperature, wind, cloudiness and precipitation) likely to arise from the predicted changes in pressure and temperature. Calculation, in the strict sense of the word, plays a relatively minor part in this process, which is essentially one of imagination disciplined by study of the physics and dynamics of the atmosphere. Conventional forecasting is as much an art as a science. If the construction of prebaratics could be made to depend entirely upon mathematics, a great step forward would have been taken. Can this be done?

---

† A familiar example of a prebaratic is the chart of ' tomorrow's weather ' shown by the meteorologist on television.

A bold attempt to establish an entirely objective method of forecasting pressure changes was made in 1922 by the late L. F. Richardson, who used finite differences and a step-by-step process of integration to solve the equations of hydrodynamics and thermodynamics in forms adapted to the atmosphere (8). The scheme was extremely ambitious in view of the meagre and inaccurate upper-air data available at the time, and prohibitively laborious with only desk computing-machines. The one forecast computed by Richardson, after months of labour, was a ludicrously inaccurate prediction of pressure changes for a period as short as six hours. But the effort was not wasted and Richardson's work may be truly said to have begun numerical weather-forecasting, even if it had to lie dormant until the high-speed electronic computer had been invented.

The initial successes with a purely mathematical method of forecasting pressure changes for short periods over relatively large areas were achieved shortly after the end of the last war with a scheme far simpler than that devised by Richardson, although the number of elementary mathematical operations involved was of the same order, and the services of very fast computing-machines had to be called upon to produce results in time to be of use. If attention is confined to large-scale systems, it is possible to ignore vertical motion and to use, not the primitive equations, but the simpler vorticity relation. The foundations of the work were laid by Rossby (9) and Sutcliffe (7).

Rossby had shown that the existence and many of the properties of the large-scale slowly moving planetary pressure waves which are found in the upper air can be deduced from a study of the change of vorticity with latitude in a perturbed zonal current, and Sutcliffe made clear the fundamental rôle of vorticity in the movement and development of weather systems. With these physical principles as a guide, the formal mathematical analysis can be very simple, as will be seen in the following section.

5. NUMERICAL FORECASTING

The equations of motion of an inviscid fluid, moving horizontally over a rotating earth are, with the notation used above,

$$\left.\begin{aligned}\frac{\partial u}{\partial t}+u\frac{\partial u}{\partial x}+v\frac{\partial u}{\partial y}&=\lambda v-\frac{\partial h}{\partial x}\\ \frac{\partial v}{\partial t}+u\frac{\partial v}{\partial x}+v\frac{\partial v}{\partial y}&=-\lambda u-\frac{\partial h}{\partial y}\end{aligned}\right\}\qquad(4)$$

The axes are fixed in the earth, and t is time.

In large-scale systems only the vertical component of vorticity, which measures the rotation of the air particles in the horizontal plane, is significant.

This component is

$$\zeta=\frac{\partial v}{\partial x}-\frac{\partial u}{\partial y}\ . \tag{5}$$

The vertical component of the earth's rotation is ω sin ϕ. In meteorology, ζ is called *relative vorticity* and η, defined by

$$\eta=\zeta+2\omega\sin\phi \tag{6}$$

is called *absolute vorticity*. From these equations it is easily shown by cross differentiation that

$$\frac{\partial\eta}{\partial t}+u\frac{\partial\eta}{\partial x}+v\frac{\partial\eta}{\partial y}=-\eta\left(\frac{\partial u}{\partial x}+\frac{\partial v}{\partial y}\right). \tag{7}$$

This is the *vorticity equation*. The right-hand side of this equation is the horizontal divergence of the wind, which vanishes for a system in strict geostrophic balance. The vorticity equation then becomes simply

$$\frac{\partial\eta}{\partial t}+u\frac{\partial\eta}{\partial x}+v\frac{\partial\eta}{\partial y}=0 \tag{8}$$

which states that in these circumstances an air particle conserves its absolute vorticity as it moves over the earth. It also follows from eqns. (1) and (5) that the vorticity of the geostrophic wind relative to the earth is proportional to the Laplacian of the contour-height, the precise relation being

$$\zeta=\frac{1}{\lambda}\left(\frac{\partial^2 h}{\partial x^2}+\frac{\partial^2 h}{\partial y^2}\right)=\frac{1}{\lambda}\nabla^2(h)\ . \tag{9}$$

It is now a simple matter to derive a *forecasting equation*. The time-rate of change of contour-height h is effectively the rate of change of barometric pressure. Simple mathematical manipulation then yields an equation for $\partial h/\partial t$, namely

$$\nabla^2\left(\frac{\partial h}{\partial t}\right)=\frac{\partial\eta}{\partial x}\frac{\partial h}{\partial y}-\frac{\partial\eta}{\partial y}\frac{\partial h}{\partial x} \tag{10}$$

and this forms the starting point for the calculation of $h(x, y)$ from a given initial distribution.

The computing scheme is relatively simple. Equation (10) is a second-order partial differential equation of the Poisson type in which the right-hand side can be regarded as a known function of h. The solution can be found over an area of the earth's surface provided that at some zero of time h is known at all points in the area and $\partial h/\partial t$ is known on the boundaries at all times. In the Meteorological Office trials a rectangular area centred on the British Isles, covering about 30 degrees of latitude and 80 degrees of longitude, is used. The values for h for, say, the 500 mb surface (about $5\frac{1}{2}$ km above sea-level) can be calculated from radio-sonde ascents with the aid of the hydrostatic equation, and a contour map can be constructed for the whole region. The trial area is then divided into about 200 sub-rectangles, of side about 300 km, and the values of h at every lattice-point are deduced by interpolation. This array of contour heights constitutes the initial data of the problem. The values of $\partial h/\partial t$ on the boundaries have to be assigned arbitrarily and thus may be considerably in error, but provided that the period of the forecast is not too long, any deleterious effects are confined to the fringes of the area. In practice, the forecast is regarded as applying only to an inner rectangle.

The forecasting equation is then solved by the machine at every one of the lattice-points by finite-difference approximations. The result is an approximate value of $\partial h/\partial t$ at the zero of time (t_0). The values of h at a short time Δt later are derived by an extrapolation formula, e.g.,

$$h(t_0 + \Delta t) \simeq h(t_0) + \left(\frac{\partial h}{\partial t}\right)_0 \Delta t.$$

The time-step Δt is usually half an hour. The result is effectively a prebaratic which is used to provide the initial data for the next step, and so on, until the whole of the forecast period is covered. The whole operation, which involves many millions of elementary mathematical operations, can be performed by a modern digital computer in a few tens of minutes. Usually, the final result is printed, by an electric typewriter attached to the machine, in the form of an horizontal array of contour-heights superimposed on an ordinary geographical map. As in conventional forecasting, the elements of weather over the area have to be deduced by a human forecaster to complete the forecast.

The mathematical process described above thus consists of the study of changes in the topography of an isobaric surface when the motion field is subject to the constraints of the geostrophic balance and the conservation of absolute vorticity. How far is this situation representative of the real atmosphere? In a typical depression there is convergence (inflow) of air in the lower levels and a slightly greater divergence (outflow) above, so that there is an intermediate ‘ level of non-divergence ’. This level is usually identified with the 500 or 600 mb surfaces, and at these heights the frictional effects of the surface are negligible. The model should therefore give the best results at about 5 km above sea-level, and this expectation had been realized. The ‘ prontours ’ of the 500 mb surface produced by this method are, on the average, about as good as, or even slightly better than, those made by experienced forecasters using time-honoured methods of extrapolation.

The model which has been described is usually referred to as *barotropic.* A barotropic atmosphere is an idealization of the real atmosphere in which it is supposed that surfaces of equal density coincide with the isobaric surfaces. There are no horizontal gradients of temperature and no change of wind with height, so that conditions on any one isobaric surface are typical of the whole atmosphere. Nevertheless, despite the obvious limitations, barotropic forecasts of pressure changes over periods of the order of 12 hours have often been surprisingly good when applied to conditions at sea-level.

Work is now in progress in many parts of the world to relax the strict conditions of the barotropic model. The real atmosphere is *baroclinic*, with well-marked horizontal gradients of temperature, and the wind at altitude is the result of the superposition of the surface pressure field on the upper thermal field indicated by the pattern of the thickness lines. Attempts have therefore been made to include the effects of temperature. The Sawyer-Bushby model, which is now being studied in the Meteorological Office with the aid of a Ferranti ‘ Mercury ’ computer, involves the solution of two simultaneous partial differential equations for the variation of contour-height and thickness, respectively. The inclusion of the geostrophic balance has the desirable effect of filtering out some complicating high-frequency ' noise ' motions, but is too restrictive in other ways, and much effort is being devoted to the study of the so-called ' balance equation ', using a stream function. It is also desirable to include topographical effects and the influences of non-adiabatic heating, such as the warming or cooling of the air by the surface over which it is passing, in the analysis.

The successful introduction of genuine, though necessarily approximate, mathematical methods into the routine preparation of prebaratics for periods up to 48 hours ahead is a landmark in the path of meteorology, and it seems certain that these methods, including the objective analysis of the intitial situation from the observations, will in time come into use in most, if not all, of the national services.

6. The problem of the general circulation

We now pass to a problem which, in its mathematical aspects, has many affinities with that described above but involves motions of much greater horizontal length-scales. This is the study of the general circulation of the atmosphere over the earth. It has been known for centuries that there is a quasi-permanent pattern of surface winds over the globe. The most prominent features of the circulation, familiar in textbooks of geography, are the sub-tropical trade wind belts of steady NE and SE winds, the broad mid-latitude belts of prevailing westerly winds and the polar easterlies. More recently, winds (observed for the most part by radar techniques) have been studied on a global basis to heights of about 25 km, a layer which contains nearly the whole of the mass of the atmosphere. From the diagrams it is evident that the mean motion of the air is predominantly zonal (parallel to lines of latitude). Westerly winds generally increase with height in the troposphere, and there are several narrow sinuous belts of very high winds situated just below the tropopause. These high-level currents, called '*jet streams*', often attain speeds of over 100 knots and are most prominent in the sub-tropical and sub-polar regions. The mean meridional motion is less easily detected, being of the magnitude of 1 metre a second, and on the scale of the general circulation, the mean vertical motion is extremely small, a matter of centimetres or millimetres a second only, and therefore cannot be measured directly.

The ' basic ' mean motion of the troposphere is a vast circumpolar whirl of westerly winds, but the pattern is more complicated and cellular in character at sea-level. The magnitude and direction of the upper vortex can be accounted for by the general fall of temperature from the equator to the poles. The easterly motion of the trade-wind belts can be explained on lines suggested by George Hadley in 1735, that is, by thermally-induced ascending motion in the tropics and a general descent of air at about latitude 30°, with the Coriolis acceleration to deflect the horizontal flow into the NE and SE trades in the lower levels. But the motion of the lower air is subject to the powerful frictional drag of the surface, and a belt of low-level winds mainly in one direction must be compensated by an opposing motion elsewhere if the angular momentum of the earth and its atmosphere is to remain sensibly constant over long periods of time. Because of friction and the rotation of the earth, there is an input of westerly momentum from the surface to the air in a belt of prevailing easterlies, and a

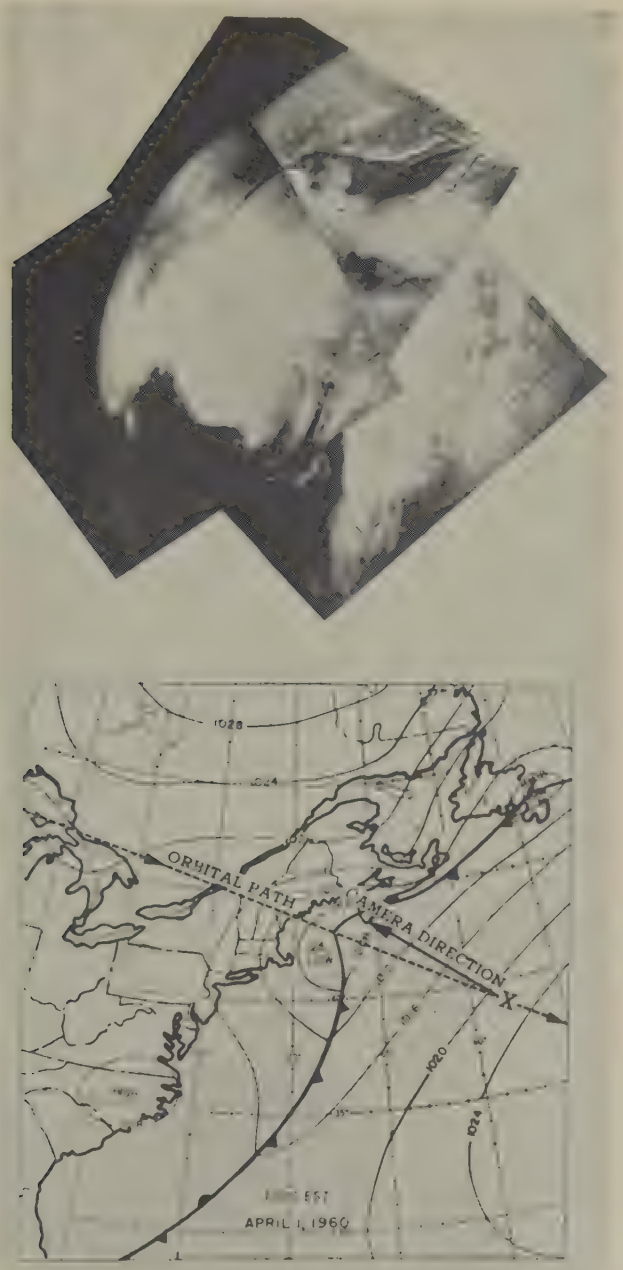


Fig. 1. The cloud system in a young cyclonic storm, located east of New England, photographed by TIROS on its third orbit the day it was launched, April 1, 1960, at 11.55 Eastern Standard Time.

A widespread cloud mass with some suggestion of a spiral flow is shown just to the right of the dark area. The dark area represents the cloudless cold air which was flowing over the middle and South Atlantic coast of the United States at that time.

A weak cold front is shown by the fibrous cloud trailing to the south out of the main cloud mass. The dark area in the northern part of the picture is the Gulf of St. Lawrence, and the St. Lawrence River.

The appropriate surface weather map is given below the photograph. The large " X " shows the geographic position from which TIROS took the photograph as it travelled from west to east at an altitude of about 450 miles. [*By courtesy of the United States Department of Commerce Weather Bureau, Washington* 25, *D.C.*]

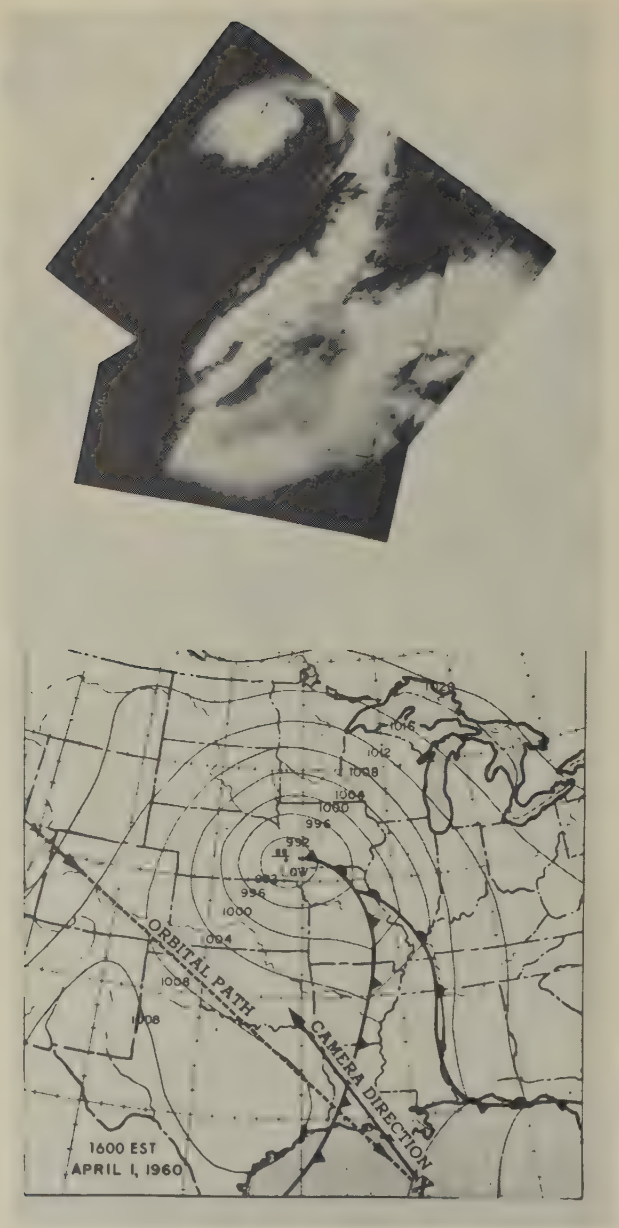


Fig. 2. A well developed storm centred over Nebraska at 3.30 p.m. E.S.T. on April 1, 1960. The photograph was taken by TIROS during its fifth orbit on the day it was launched.

The centre of the storm is revealed by the circular cloud mass in the upper part of the picture. A prominent cold front is seen trailing from the storm centre to the south and southwest. The large cloud mass in the foreground was located near the Gulf Coast of Louisiana and Mississippi.

The appropriate surface weather map is given below the photograph. The large " X " shows the position from which TIROS took the photograph.

[*By courtesy of the United States Department of Commerce Weather Bureau, Washington* 25, *D.C.*]

flow of momentum in the opposite direction in prevailing westerlies. Without a continuous supply of angular momentum, the surface westerlies would be reduced to a calm by friction in about 10 days.

The study of angular-momentum exchange over the surface of the earth was initiated by Sir Harold Jeffreys who, over 30 years ago, suggested that the polewards transfer from the tropical regions is mainly accomplished by large-scale turbulence, with the mid-latitude cyclones and anti-cyclones playing the part of eddies. According to Jeffreys, the exchange process involves a mechanism like that found in the lowest levels of the atmosphere and studied in micro-meteorology (see section 2 above) but on a much larger scale. This theory remained for many years of academic interest only, but after the last war it became possible, with the wealth of upper-air data available, to put it to the test. Actually, two mechanisms are at work; transfer by mean meridional motion and the large-scale *austausch* process. The studies have shown fairly conclusively that in the equation of transfer the meridional motion or ‘ drift ’ effect is significant only in the tropics and sub-tropics, and that elsewhere the eddy-flux term suggested by Jeffreys is dominant. Thus the tropics and sub-tropics are the main sources of angular momentum and the mid-latitudes the main sink, with the large-scale disturbances acting as the principal agents of transfer.

The internal consistency of these results has created among meteorologists a feeling of confidence that the more important parts of the atmospheric engine have now been positively identified and their action explained quantitatively in a tolerably satisfactory manner. But many of the details of the circulation are still obscure, and more sophisticated mathematical theories are required to deal with them. Among the most interesting and promising work of recent years is the approach by means of mathematical ‘ experiments ’ based on baroclinic theories and carried out with high-speed computers. Ideally, the problem of the general circulation is to produce a complete and consistent physical description of the genesis and maintenance of the observed pressure, temperature and motion fields of the atmosphere starting from the solar constant and the configuration of the sun-earth system. At present, meteorologists are content to limit their studies to a less ambitious problem, to account for the mean motion when the average temperature distribution over the earth is regarded as given.

The initial attempt on this problem was by N. A. Phillips (10), who began with a dry atmosphere at rest relative to a rotating earth for which temperature is a simple function of latitude. Frictional effects were included by semi-empirical terms in the equations of motion, which were then integrated by finite-difference methods using a digital computer. The initial motion-field, which was established in an interval corresponding to some weeks of natural time, was of westerly zonal winds with a weak meridional component. The isobaric surfaces were then perturbed in a random fashion (as would inevitably happen in the real atmosphere because of the effects of topography and local heating) and, after about a month of real time, the motion developed wave-form patterns which bore a strong resemblance to the observed circulation, with an easterly-westerly-easterly distribution of zonal flow between the equator and the poles. The work has been continued by J. Smagorinsky (11) who, using the primitive equations of motion, has succeeded in producing an even closer approximation to reality. In this experiment, incoming radiation is made a simple function of latitude and energy is dispersed by black-body radiation from the earth, surface friction, and internal convective stresses. The lateral exchange of heat and momentum is mainly by an *austausch* process. After integration covering a period of about 60 days of natural time, the resulting motion fields exhibited a northwards transport of momentum and heat which agreed quantitatively with the fluxes deduced from observations of the real circulation.

The word ' experiment ' has been used advisedly in the above account. The complexity of the physical problem and the non-linearity of the basic equations means that no formal solution can be found. Instead, the mathematician, like the experimental physicist seeking to imitate the actual circulation with rotating bowls of water heated at the circumference and cooled at the centre, tries to reproduce the circumstances of atmospheric phenomena by investigating the consequences of certain simple distributions of energy sources and sinks in mathematically-specified model atmospheres. The encouraging (and somewhat unexpected) result is that even fairly crude models behave very much like the real atmosphere, and in many instances the agreement is quantitative. As little as 20 or 30 years ago few meteorologists would have believed this to be possible.

The analysis of the general circulation by these means has renewed hopes that someday it may be possible to effect a break-through in the notorious problem of long-range forecasting. At present there is no way known to meteorologists of foreseeing marked anomalies in climate, such as exceptionally dry summers or severe winters, many months in advance. No real success has attended the search for weather cycles or the connexion between extra-terrestrial phenomena (such as sunspots) and the weather. If, however, a trustworthy theory of the general circulation can be developed—if, in other words, meteorologists can account quantitatively for the average state—it may be possible also to predict the consequences of specified perturbations of this state. At present, the deployment of scientific staff in long-range forecasting research is in some ways an act of faith, for there is no assurance that it will ever be possible to predict weather more than a few days ahead. But if even somewhat imprecise but reliable statements about the general character of a coming season could be made well in advance, the economic results would be tremendous, affecting the lives of everyone, and especially those who live in the underdeveloped areas of the world, where a prolonged drought still holds out the threat of famine.

7. New methods of observation

Meteorologists have long suffered from their inability to look at the atmosphere from above. This difficulty is now being overcome, and two of the more striking achievements of ' the entry into the space age ' are the successes of the American radiation-measuring and cloud-viewing satellites.

The heat balance between the atmosphere and the sun is fundamental in theoretical meteorology, but until now the details of the balance sheet have had to be filled-in from climatological data. A satellite experiment to measure directly the gross short- and long-wave radiation entering and leaving the atmosphere was first proposed by V. E. Suomi, and is now in progress. The preliminary reports are encouraging.

The importance of a successful Suomi-type experiment is undeniable. A regular series of measurements of radiation exchange made from outside the atmosphere should throw considerable light on possible variations in the energy budget, and it is not impossible that such relatively short-period fluctuations, if they exist and can be located and measured to the required degree of accuracy, may prove to be some of the missing links in the unsolved problem of long-range forecasting. With the present satellite, it is possible to follow the movements of high and low infra-red radiation loss with weather systems, and overall measurements of emission agree well with data obtained from aircraft. With more highly developed instruments, the potentialities of the method are very great.

The Suomi experiment resembles the use of an exposure-meter in photography. The *TIROS* cloud-viewing satellite is a camera which records the visible detail but gives no quantitative information concerning the radiation flux. This remarkable instrument circles the earth in about 100 minutes at a height of about 450 miles, and its main task is to produce pictures of high-level clouds. It incorporates two lenses, one with a field of 104° and the other of 12°. The image of the sunlit earth is retained for about 10 seconds on a 1½-in. diameter television tube and is scanned by 500 lines. Pictures of areas about 750 and 65 miles square, respectively, have been successfully received. During its working life *TIROS* produced about 24 000 pictures, of which some two-thirds are believed to contain information of importance in meteorology.

The *TIROS* pictures (figs. 1 and 2) are not always easy to interpret, for it is at times almost impossible to distinguish between sheets of cloud and areas of snow-covered landscape. The most striking evidence of the potentialities of this means of observing upper winds comes from pictures of the cloud systems associated with relatively compact disturbances, such as hurricanes, in poorly observed areas such as the Indian and Pacific Oceans. Such information could be of considerable assistance to the short-range forecaster, for when the satellite passes over the receiving station, a picture can be received and processed for transmission to a weather office in about 15 minutes.

The successful establishment of a global system of reception of cloud pictures from satellites does not necessarily mean that all the forecasting services of the world will take a great stride forward. In areas well covered by a network of observing stations, the information from the satellites may add little to what is already known. On the other hand, there is no doubt that the observations will help in the problem of the general circulation. Satellite pictures and high-speed computer experiments could, and no doubt someday will, form a coherent whole. When this is achieved, the advance of meteorology toward the goal of an exact science will be greatly accelerated.

References

(1) Sutton, O. G., 1953, *Micrometeorology*, McGraw-Hill, New York.
(2) Taylor, G. I., 1922, ‘ Diffusion by Continuous Movement ’, *Proc. London Math. Soc.*, **20,** 196–212.
(3) Kolmogoroff, A. N., 1941, *C.R. Acad. Sci. U.R.S.S.*, **30,** 301.
(4) Taylor, G. I., 1915, ‘ Eddy Motion in the Atmosphere ’, *Phil. Trans. Roy. Soc.*, A, **215,** 1–26.
(5) F. N. Frenkiel and P. A. Sheppard (editors), 1959, *Atmospheric Diffusion and Air Pollution.* Advances in Geophysics, 6.
(6) Sutton, O. G., 1960, *Understanding Weather*, Pelican Books, London.
(7) Sutcliffe, R. C., 1947, ‘A Contribution to the Problem of Development ’. *Quart. J. Roy. Met. Soc.*, **73,** 370–383.

(8) Richardson, L. F., 1922, *Weather Prediction by Numerical Process*, Cambridge Univ. Press.
(9) Rossey, C.-G., 1940, 'Planetary Flow Patterns in the Atmosphere'. *Quart. J. Roy. Met. Soc.*, **66,** (Supp.) 68–87.
(10) Phillips, N. A., 1956, 'The general circulation of the atmosphere; a numerical experiment', *Quart. J. Roy. Met. Soc.*, **82,** 123–164.
(11) Smagorinsky, J., 1960, Programme of the International Association of Meteorology and Atmospheric Physics, Helsinki (abstract only).

The Author:

Sir Graham Sutton, C.B.E., D.SC., F.R.S., is Director-General of the Meteorological Office. He was educated at Pontywaun Grammar School, The University College of Wales, Aberystwyth and Jesus College, Oxford (of which he is now an Honorary Fellow). Most of his life has been spent in government scientific establishments, and he has worked in many fields, including meteorology, rocket aerodynamics, chemical defence, radar and tank armaments. From 1947 to 1953 he was Bashforth Professor of Mathematical Physics and later Dean of the Royal Military College of Science, Shrivenham. He has published several books, on meteorology, aerodynamics and applied mathematics.

X-ray Microscopy

by W. C. NIXON
Engineering Laboratory, Cambridge University

SUMMARY

The wavelength limitation on resolution indicates that an X-ray microscope may surpass the optical microscope but will not compete with the electron microscope even if all the technical difficulties could be overcome. However, there are other advantages besides resolution that make X-rays an attractive medium for microscopy.

The penetration of X-rays allows the examination of the internal detail of a specimen on the micro scale without the need for sectioning and reconstruction. The large depth of focus permits the viewing of this internal detail with equal sharpness and stereographically if necessary. The specimen may be in air at atmospheric pressure, while the electron microscope demands high vacuum, and yet the resolution may be better than that of the optical microscope. Most important, the simple absorption and emission spectra of X-rays leads to the extraction of quantitative information from the image in terms of mass and elements present that in many cases cannot be obtained in any other way.

These and other reasons have led to the search for a usable X-ray microscope and the three more successful methods of reflection, contact and projection are discussed here. In addition, recent results on microanalysis show where the bulk of the future work will lie.

---

1. REFLECTION X-RAY MICROSCOPY

The attraction of X-rays as an imaging radiation for microscopy has led to consideration of both refracting and reflecting X-ray systems directly analogous to optical systems for visible light. A transmission X-ray lens is not feasible since the path length for appreciable refraction is sufficient to absorb almost completely the soft X-rays necessary for microscopy. A reflecting system with the X-rays striking a mirror at near normal incidence is also impracticable due to the penetration of the rays into the surface. However, by using a sufficiently small glancing angle of incidence the X-rays will be totally reflected back into the first medium, since the mirror's refractive index, although close to unity, is very slightly less than this figure. A cylindrical surface will focus the rays diverging from a small source into a line. By combining two of these X-ray mirrors the line image from the first serves as the line object of the second which is in turn focused into a point (fig. 1). This simple pair of crossed cylindrical sections has been studied extensively for use as an X-ray microscope. A resolution of 0·1 micron has been predicted, equal to the ultra-violet microscope, and a resolution of 1 micron achieved in practice.

The small glancing angle introduces several severe geometrical aberrations, some of which might be eliminated by the use of further correcting mirror pairs. The reflected intensity is quite low even for the first mirror and the exposure time reached experimentally is about half-an-hour. Other reflecting elements would prolong this time so that the alignment procedure becomes impossible to carry out. In addition to the normal types of aberration, such as spherical, coma, and curvature of field, the placing of the mirrors one after the other leads to a difference in magnification in two planes of the image at right angles. A further difficulty in use is the need for extremely smooth surfaces over the whole area of the mirror. Ideally, the surface finish should be much better than the resolution sought, say no roughness above a few hundred angstroms over at least one square centimetre. These two difficulties, surface shape and surface roughness, have both limited the practical resolution to about one micron.

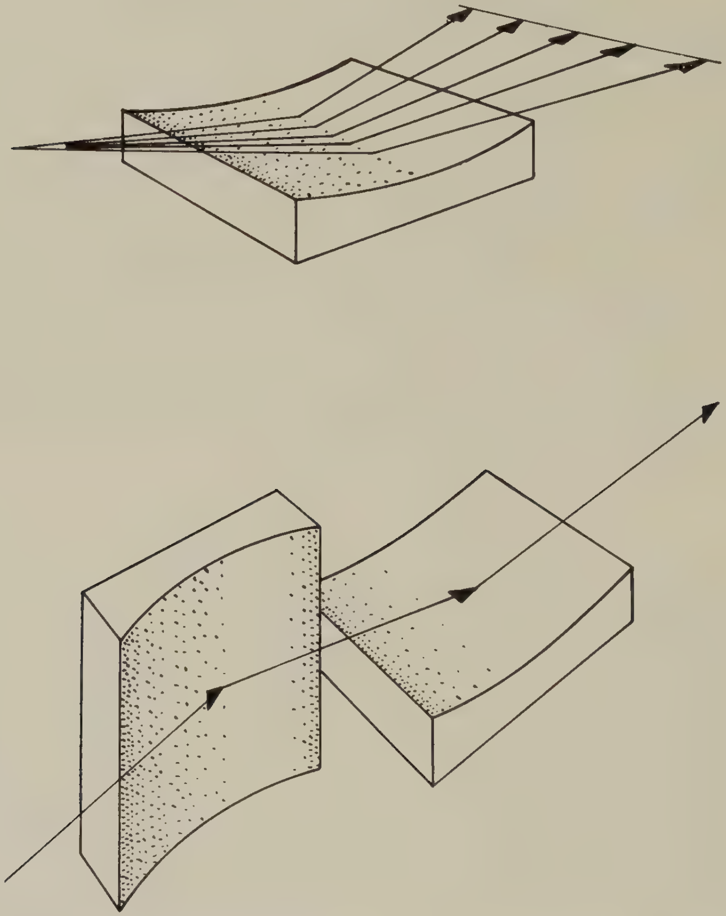


Fig. 1. An x-ray mirror focusing a point into a line (above) and when used as a crossed pair, focusing a point into a point (below) (after P. Kirkpatrick).

If new techniques overcome these two limitations the wave nature of the radiation used will set a more fundamental limit to the resolution. As with an optical microscope the Fresnel diffraction of the rays at the specimen gives rise to an edge diffraction pattern composed of fringes that will obscure the geometrical images of two objects that are close together. The approximate relationship between the resolution d, the wavelength λ and the angular aperture α is $d=\lambda/\alpha$. Using visible light it is possible to produce objectives with a very large angular aperture so that the resolution may be at best about one-half the wavelength used, say 0·2 micron for blue light. With x-rays the wavelength is some

400 times shorter, say 10 Å, but the angular aperture is also reduced because of the need for a small glancing angle. For $\alpha = 10^{-2}$ we see that the x-ray resolution will be 1000 Å (0·1 micron) on fundamental grounds alone, irrespective of the perfection of the optical system. The glancing angle may be slightly larger in some favourable cases but the broad conclusion still stands: the reflection x-ray microscope is difficult to produce even at the moderate resolution of 1 micron and tremendous barriers must be surmounted for even a modest improvement on this figure.

2. CONTACT MICRORADIOGRAPHY

Despite the drawbacks preventing the development of the reflection x-ray microscope it has been possible to obtain magnified images using x-rays as the initial imaging medium. A normal x-ray tube as shown on the left of fig. 2 illuminates the specimen which is placed in very close contact with a fine grained photographic emulsion. The x-ray image is recorded at unit magnification and after development the negative is magnified by an optical microscope. This simple procedure has been used extensively for microradiography of both biological and metallurgical specimens since fine grained film became generally available about 25 years ago. Refinements in technique have reduced the practical difficulties so that the full limit of resolution of the optical microscope, about 0·2 micron, may be used.

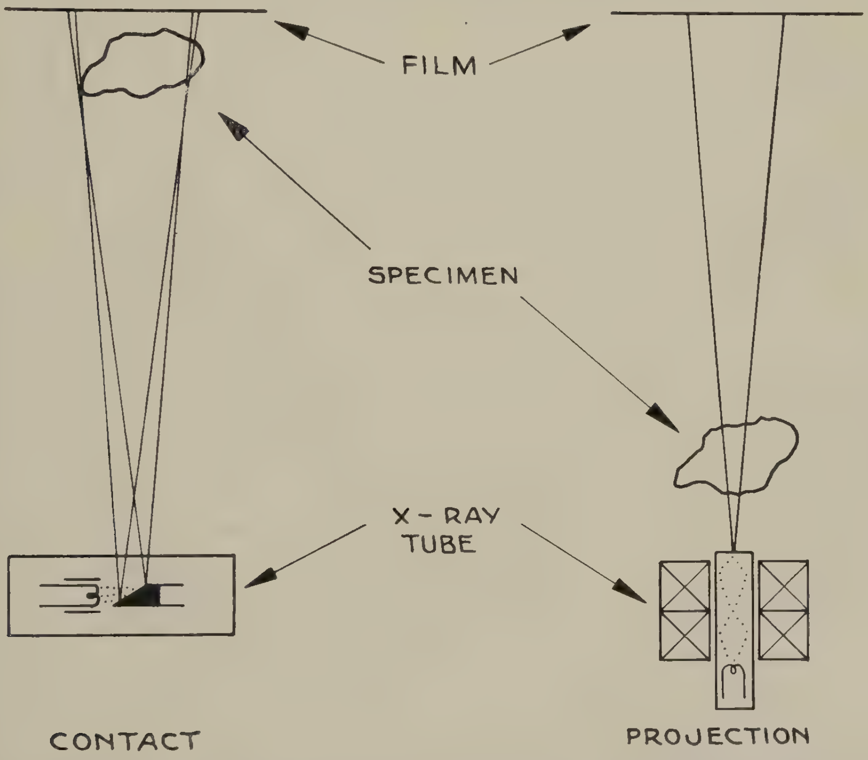


Fig. 2. Contact microradiography (left) using a normal x-ray tube and fine-grained photographic film. Projection x-ray microscopy (right) using a special x-ray tube and normal film (after V. E. Cosslett and W. C. Nixon).

Geometrical unsharpness of the image depends on the size of the focal spot in the x-ray tube, the distance between the tube and film, and the distance between the lower surface of the specimen and film. A 200 micron section placed at 1 metre from a 1 mm focal spot would be blurred by the amount stated above, 0·2 micron. This is a relatively thin section and yet is thicker than the usual 5 to 10 micron histological section prepared for optical microscopy by normal fixation and staining methods. The 200 micron section is equally sharp in depth when recorded on the film, which in turn is only a few microns in thickness and so lies within the depth of field of the optical microscope used for magnification. The penetration of the x-rays permits recording of information about the specimen in depth without serial sectioning.

The photographic grain size is well below this limit of resolution but on development grain clusters may grow to this level and impair the image unless care is taken in the processing. These fine grained films are very insensitive and a normal exposure would last from one to two hours or more at 1 metre from the x-ray tube. However, at this distance a large area of the film is uniformly illuminated and so many specimens may be radiographed at once, followed by extensive searching of the negative under the optical microscope. Recent experiments on detaching the emulsion from the supporting glass for examination in the electron microscope hold out promise of higher resolution in the future. However, there is still an x-ray wavelength limitation on resolution due to the formation of Fresnel fringes as with the reflecting x-ray microscope.

In this case the fringes are formed between the direct wave and the scattered wave arriving at the film situated a distance b from the object point. If the scattered wave makes a small angle α with respect to the initial ray, and the first fringe maximum is given by d, then $\alpha = d/b$. However, the extra path length of the scattered wave for constructive interference is λ, so by similar triangles $d = \lambda/\alpha$. Combining these equations we find that the first Fresnel fringe width, and thus the resolution limit, is given by $d = (b\lambda)^{1/2}$. For an x-ray wavelength of 10 Å and the specimen thickness of 200 microns given above (the distance b), the resolution limit will be 0·45 micron, or somewhat worse than the optical microscope. Because of the square root relationship the section thickness must be reduced to 10 microns before the resolution falls to 1000 Å as discussed for the reflection microscope. We see that even if it is possible to magnify the x-ray image with the electron microscope only very thin sections may be used due to this fundamental limit.

Despite these difficulties involving future improvements, the method has been used with advantage at the present resolution. This type of information is difficult to obtain by any other means from the interior of an optically opaque specimen. The absorption of the x-rays when passing throught the object follows the well known exponential law. The exponent in the logarithmic term contains the mass absorption coefficient of the element, the density and the specimen thickness. If any two are known then the other may be found by microphotometering the x-ray negative to compare the x-ray intensities through the specimen at two points. Usually the thickness is accurately measured and with biological material the elements are assumed to be mainly carbon, nitrogen and oxygen. Then the density, or mass per unit area of a specimen of that thickness may be found, with a spatial resolution of about 1 micron. Conversely, by using the sudden change in absorption coefficient at an absorption edge, specific elements may be identified of the higher atomic numbers, but not at present in the C, N and O region.

Contact microradiography is a simple and inexpensive additional method to be applied in conjunction with normal optical microscopy when more quantitative information is sought.

3. PROJECTION X-RAY MICROSCOPY

In this method of x-ray microscopy, shown on the right in fig. 2, an initial x-ray image is formed with some degree of magnification without the use of x-ray lenses or mirrors by using a special x-ray tube consisting of an electron gun and two lenses. The projection geometry of the system gives a penumbral unsharpness in the image approximately equal to the size of the small x-ray source when the magnification is 10 times greater. Two magnetic electron lenses will reduce the size of the initial electron source in the electron gun from 100 microns to 1 micron or less. This fine electron probe, accelerated by 10 kv strikes a thin metal foil target, which may be 3 microns of copper or 1 micron of tungsten. The x-rays are generated within a small volume of 0·5 micron diameter, depending on the kilovoltage used. The thin target is also the vacuum seal of the electron tube and the x-rays emerging from the foil in the direction of the electron beam may be used to form an enlarged shadow projection image. The magnification is given by the ratio of the film distance to specimen distance. Since the foil is only a few microns thick and one side is at atmospheric pressure it is possible to place the specimen within 10 microns of the x-ray source. A film distance of 10 mm then gives an x-ray magnification of 1000 × before any further optical magnification. This initial x-ray enlargement justifies the term x-ray microscope for this method of microscopy. A more normal specimen distance is 100 to 500 microns leading to a magnification of 100 × at 1 to 5 cm film distance, followed by 10 × or more photographic enlargement. The useful cone of x-rays fills an angle of 60° without appreciable geometrical distortion and so the larger specimen distance leads to a larger field of view on each negative.

The same Fresnel diffraction limitation on the ultimate resolution may be calculated for this case as above. The specimen distance, b, is directly analogous to that given in the contact microradiographic method, so that $d=(b\lambda)^{1/2}$ as before. Again, with the specimen 10 microns from the target and λ of 10 Å the fringe width is equal to 0·1 micron. This size of x-ray Fresnel fringe has been demonstrated under these conditions when the size of the x-ray source itself was no greater than this figure. It is worth noting that this small specimen distance is not possible with the normal type of solid anode x-ray tube and that high magnification projection x-ray microscopy must be carried out with a thin window target. Also, this same Fresnel limit rules out the possibility of high magnification optical micrographs by projection since the wavelengths of visible light are very much longer. Even if the optical source were small enough, say 1 micron, and the specimen could be placed 10 microns away, then the resolution for blue light at 4000 Å would be 2 microns or some 10 times *worse* than that achieved with the best optical microscope of the transmission refracting type.

Further advances in resolution are limited by the electron optical system as well as by diffraction effects. In any optical arrangement the sine law predicts the relative brightness of the object and image without regard to the ray paths.

With the given electron sources and severely aberrated electron lenses available at the moment, the electron intensity into the focused spot varies as the cube of the spot size. In other words, if the electron spot is reduced from 1 micron to 0·1 micron the x-ray exposure time will rise by 1000 times, all other factors being equal. This is approximately the increase found in practice, from a few seconds at 1 micron resolution to 10 minutes or more even at a smaller film distance, when the source size approaches 0·1 micron. Disregarding the Fresnel limit, this intensity difficulty would predict an exposure time of several days if a source of 100 Å were produced, or well beyond the mechanical and electrical stability of the system. As an aside it is interesting to note that the heat dissipation in the thin metal foil target increases as the size is reduced. The total load is becoming smaller but the specific load will rise well above that found in a normal x-ray tube. The heat can escape more easily from the centre of a small focal spot and so water cooling is not necessary for spots below about 10 microns.

At the present resolution this method of x-ray microscopy may be used to produce images of thicker specimens than the contact method and with the same quantitative information given by the variation in density in the x-ray negative. Two examples are given in figs. 3 and 4. The spider leg shown in fig. 3 has been dried in air but had not shrunk appreciably. The dark muscle in the centre of the limb is more x-ray absorbing than the outer hard shell. A few hairs may be seen on this surface and in some cases these are seen to be hollow. No dissection is necessary to see the interior of the specimen and a stereographic

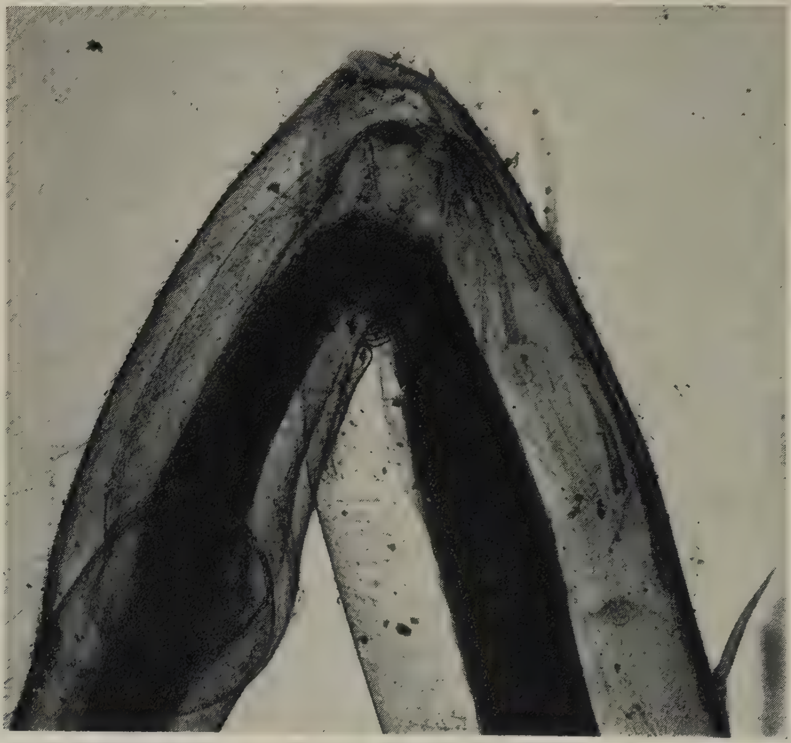


Fig. 3. Projection x-ray micrograph of a spider leg. (W. C. Nixon and J. W. S. Pringle.)

view is possible by taking two successive exposures with slight alteration in the specimen position between. A metallurgical example is shown in fig. 4 where the specimen is 500 microns thick and is composed of aluminium with 5 per cent tin. The tin is more X-ray absorbing and appears along the grain boundaries and also within the grain in small particles. A few white cracks are seen at the grain boundaries where the tin has been fractured. The stereographic view of this specimen is more revealing and has been used in a study of the angles between grain boundaries where three grains meet. Since the specimen is not in contact with the film or the X-ray tube it is possible in principle to carry out dynamic tests such as straining, fatigue, fracture, heating and cooling while viewing the magnified X-ray image. Many more biological, metallurgical and other types of application of X-ray microscopy will be found in the general references listed at the end of this article.

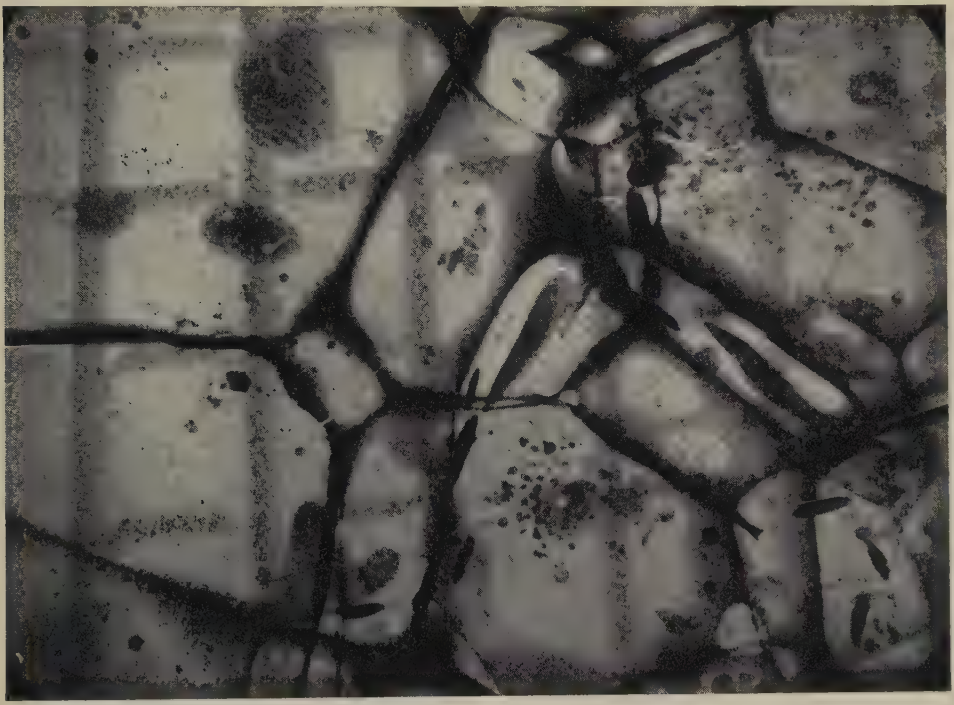


Fig. 4. Projection X-ray micrograph of an aluminium—5 per cent tin alloy, with the tin showing dark along the grain boundaries. (W. C. Nixon, *Proceedings of the London Conference on Electron Microscopy*, 1954; p. 307, Royal Microscopical Society, 1956.)

4. X-RAY MICROANALYSIS WITH THE SPECIMEN IN AIR

In the previous methods of X-ray microscopy a small selected area is analysed by scanning the X-ray negative with a microphotometer to determine the relative absorption of various parts of the specimen. With modern X-ray counting techniques it is possible to eliminate the photographic step and analyse a small region of the specimen directly, as shown in fig. 5. The electron beam from an electron gun is focused to a small source on a metal foil target exactly as in the projection x-ray microscope. The specimen at distance t away casts an enlarged x-ray shadow image that in this case falls on to an opaque screen containing a small aperture of diameter d at a distance D from the target. The x-rays that have passed through this aperture have also passed through a smaller area of the specimen, of diameter td/D. In one example with $d = 3$ mm, $t = 0{\cdot}5$ mm and $D = 150$ mm the area of the specimen analyzed was 10 microns. This small area is selected under the optical microscope and is then placed on the axis of the x-ray cone by reference to the fluorescent screen image of the specimen.

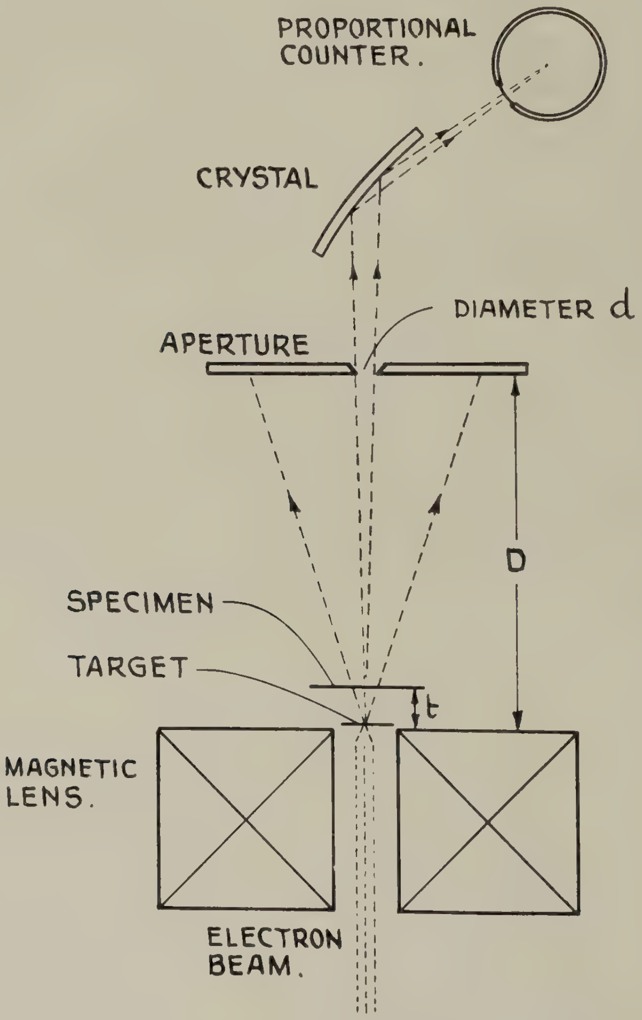


Fig. 5. Point focus x-ray tube used for microanalysis. (J. V. P. Long and J. D. C. McConnell, *Mineralogical Magazine*, 32, 117, 1959.)

When the fluorescent screen is removed the x-rays pass through the aperture and are analyzed by the crystal spectrometer and proportional counter. The x-ray tube is run at 25 kv and 10 microamps on a 2 micron spot or a specific loading of 80 kw/mm^2. The relatively low x-ray intensity in a small band of wavelengths determines the counting time which can be up to half-an-hour for a few per cent accuracy. This electronic method of absorption microanalysis gives a direct result without the difficulties of calibrating the photographic plate and standardizing the processing procedure.

If the aperture is made much smaller and placed below the specimen then a very narrow beam of x-rays, 5 to 10 microns in diameter, will pass through an equally small region of the specimen and the fluorescent x-rays excited by the primary beam may be detected at one side. With the absorption method the relatively high background limits the ultimate sensitivity in terms of the detection of minute concentrations within the small area. The background in the fluorescence method is very much lower leading to a sensitivity of a few parts in 10^5, which is much higher than with the absorption method.

With a crystalline specimen the beam of x-rays will be diffracted and this method of analysis may also be carried out on small area with the specimen in air at atmospheric pressure. The same geometry as for fluorescence analysis is used for forward diffraction by transmission and the high specific loading of the x-ray tube leads to short exposure times. A 10 micron area of silver grains gave a spotty ring diffraction pattern in ten minutes while the same area required ten hours using a sealed-off x-ray tube with a large focal spot and a small lead glass capillary collimator. When the aperture is removed the divergent beam of x-rays will produce Kossel line diffraction in single crystals and this method is very sensitive to the state of perfection of the crystal. The diffraction resolution depends on the size of the x-ray source and as this is 1 micron the Kossel line diagrams may be enlarged several hundred times for high accuracy measurements. After straining and recrystallization the lines are broken up into short arcs which may be used to follow the state of deformation of the specimen.

5. X-RAY MICROANALYSIS BY DIRECT EMISSION

If the specimen may be placed in the same vacuum that is necessary for the electron beam then the specimen and target may be identical. The characteristic x-rays generated at the point of impact of the electron beam are used for analysis rather than for microscopy as discussed in the previous sections. The electron beam will penetrate a small fraction of a micron into material of medium density when the accelerating voltage is about 10 kv. This means that it is essentially a surface analysis and that one surface only of a bulk specimen need be prepared as in normal metallography. The K or L x-rays that are emitted are analysed by a crystal spectrometer (quartz and lithium fluoride curved crystals) and/or proportional counter (argon + a few per cent CO_2) in the same manner as in microscopic x-ray fluorescence analysis. The key problems are locating the area struck by the electron beam and ensuring sufficient x-ray intensity so that high accuracy is achieved for small concentrations over a short recording time. An optical microscope with a long working distance objective at one side of the electron beam has been used to view the exact region analysed during the actual recording of the emitted x-rays. With most vacuum systems a deposit of carbon contamination is built up where the electron beam has struck and the localized area may be verified in this way after the analysis. The cross-hairs of the microscope are calibrated to coincide with the electron beam by this means initially, but a magnetic specimen may deflect the beam by a different amount at different regions on the surface. The resolution is limited partly by the long working distance employed but may be improved somewhat by using reflecting objectives that are concentric with the electron beam. In addition, the specimen is viewed with normal incidence giving a flatter field of view. Alternatively, the specimen may be viewed at high optical magnification and subsequently turned into the path of the electron beam for analysis by rotation of a shaft or a drum. This can be set repetitively to one micron, or equal to the average spot size used in most applications. The contamination spot still gives an indication of where the analysis has been made after the event.

Another approach to localization of the analysed region is to use the technique of scanning electron microscopy and display the surface detail of the specimen on a cathode ray tube long persistence screen scanned in synchronism with the electron beam in the column of the x-ray microanalyser. The resolution is now given by the size of the electron beam and therefore is equal to the analysed area, which is slightly worse than the best optical resolution in most cases. One particular advantage of this method is the display of the surface distribution of a particular element by using a selected x-ray emission line to modulate the cathode ray tube instead of the signal from the scattered or secondary electrons. In practice two cathode ray tubes are used, one continuously presenting the electron image and the other the x-ray image. Indeed, since all of the characteristic x-ray emission is produced at the same time and a portion only is selected it is possible in principle to duplicate the detection system and display simultaneously as many x-ray images as there are elements in the specimen surface.

The sequence of events followed in forming such a scanning image is shown in fig. 6. The electron gun and two magnetic lenses are similar to those discussed above for the point projection x-ray microscope. The total electron beam reduction is shared between the two lenses with the final lens operating at a short focal length and low spherical aberration, but with a reasonable working distance for ½ in. diameter samples. The double deflection scanning coils are inserted between the two lenses and ensure that the electron beam passes through the centre of the final lens even for large deflection angles at low magnification.

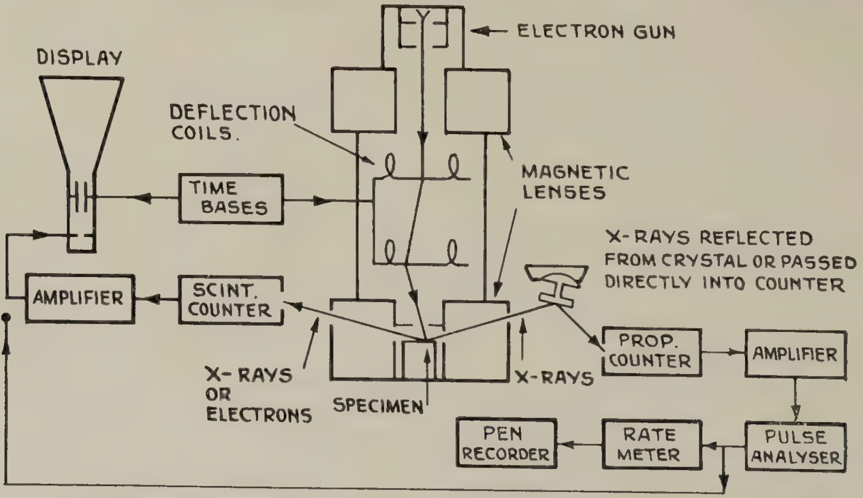


Fig. 6. Block diagram of the x-ray scanning microanalyser. (P. Duncumb, *British Journal of Applied Physics*, **10**, 420, 1959.)

The scattered electrons or the total x-rays emitted are collected by a phosphor and a photomultiplier forming the scintillation counter. The amplified signal is applied to the grid of the cathode ray tube that is scanned in parallel with the main electron beam. The magnification is increased by reducing the size of the scanning raster on the specimen surface while leaving the scan on the display tube at a constant size. This scan may be varied also, when some detail is to be examined at a higher scan rate but with a smaller field of view at the same magnification.

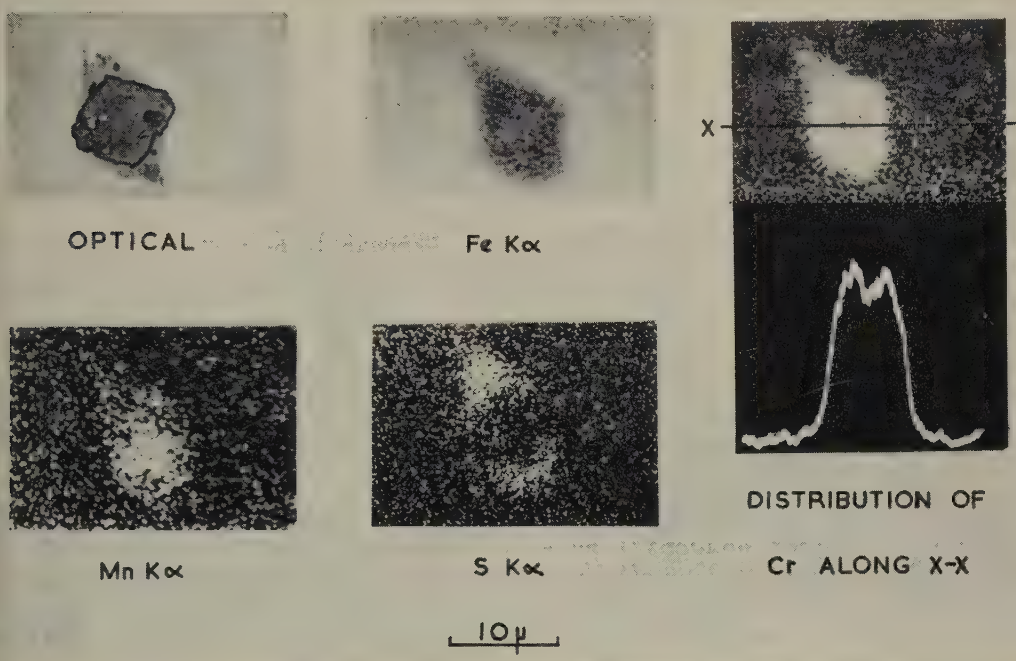


Fig. 7. X-ray scanning images in iron, chromium, manganese and sulphur together with an optical micrograph of an inclusion in steel. The chromium distribution across the centre of the inclusion is also plotted to show the dip in the middle. The magnification is given by the 10 micron mark (1/100 mm). (D. Melford and P. Duncumb, Tube Investments, Hinxton Hall, Essex.)

The crystal spectrometer and proportional counter select the characteristic x-rays from each element in turn. The crystal may be set to pass Cu K x-rays at 1·54 Å for example and the display picture formed with this signal modulating the grid of the tube and so presenting a view of the surface of the specimen in terms of the copper distribution. Alternatively, an unknown inclusion may be selected from the electron or total x-ray picture and the one micron electron probe stopped on that spot. The crystal spectrometer is then rotated through a range of Bragg angles in synchronism with a pen recorder indicating the relative amounts of each element present in the one cubic micron bombarded.

The type of result obtained by photographing the screen of the cathode ray tube is shown in fig. 7. The specimen is an inclusion (about 10 μ cube) in the surface of a steel containing 3·67 per cent chromium and other trace elements.

The optical and electron images are followed by scanning X-ray images in iron chromium, manganese and sulphur radiation showing the distribution of these elements in a qualitative way. More quantitative results were obtained by stopping the electron probe and recording the concentrations of these elements. At the outer edge of the halo surrounding the inclusion the sulphur and chromium were both over 40 per cent while the iron and manganese were both under 10 per cent; at the edge of the inclusion, the iron has risen to 14 per cent and the chromium fallen to 39 per cent while the manganese represents only 4 per cent. In the centre of the particle the iron content is approximately the same but the chromium has fallen to 36 per cent and the manganese risen to 10 per cent of the total. The depletion of chromium is shown graphically in the figure by stopping the scan in one direction and tracing the distribution of chromium along the horizontal line across the centre of the inclusion. The dip in the concentration occurs at the centre of the image. This type of information is of great help in deciding on the origin of inclusions of this type that sometimes lead to cracking and failure of the material during rolling or subsequent working.

The total collected X-ray intensity falls rapidly as the resolution is improved in this microanalyser and the present limit is reached at about $\frac{1}{4}$ micron spot size at 6 kv accelerating voltage for a scanning image of 100 lines. The original electron beam current focused into the electron probe is limited by well-known optical rules affecting the relative brightness of object and image in a lens system. Using the presently available electron guns and lenses it is possible to focus one microampere into a one micron spot at 10 kv, in round figures. This represents 6×10^{12} electrons per second arriving at the specimen. About 0·1 per cent of these electrons produce characteristic X-rays. The detection systems used until recently have a very small solid angle of collection and so pass only 0·01 per cent of the total X-rays produced. A further loss of a factor of ten occurs from the low reflexion efficiency of the crystal and detection by the counter. After these reductions of the number of quanta available for amplification there are some 6×10^4 pulses per second appearing on the screen of the tube. However, with an image of 100 lines or 100×100 picture points this represents only 6 quanta per picture point. In practice the integration time of the long-persistence screen is about 5 seconds so there are effectively 30 pulses per picture point for visual observation and 720 per picture point on the photographic record after a 2 minute exposure. For a low contrast specimen there should be at least 10^3 quanta per picture point for a noise-free image and in fact most of the scanning results do show a background distribution due to noise from the insufficient number of recorded X-rays over the time available. Information may be extracted from such noisy pictures at the present resolution but since this limit has already been reached there is little hope of reducing the resolution in a scanning X-ray image below that at present without radical changes in the electron beam system.

6. Electron microscopy and X-ray emission microanalysis

With suitable electron-transparent specimens it should be possible to combine the high resolution of the electron microscope with the precise analysis of X-ray emission. Thin metal alloy foils, extraction replicas of segregations or inclusions in a metal surface and stained biological sections could all be analysed on a point to point basis using an electron probe of a few hundred angstroms.

The useful limit of x-ray intensity is uncertain due to such factors as specimen heating, variations in atomic number and kilovoltage and efficiency of x-ray collection. However, it has been possible to achieve a resolution close to 2000 Å with an x-ray scanning system using a field of view of 100 lines and a 2 minute photographic record as discussed above. This represents the present limit of the scanning system for x-ray image display using existing electron sources and lenses. The collected x-ray intensity varies approximately as the fourth power of the resolution sought and so an improvement to 200 Å would involve the impossibly long exposure time of 2 minutes × 10^4 or 14 days. If the feature to be analysed can be located by some other means, then the electron probe may remain stationary and instead of 100 × 100 or 10^4 picture points recorded in 2 minutes, there would be only one spot irradiated and the collected intensity could be 10^4 as great for the same exposure time. This is the factor quoted above for an improvement of 10 × in analysed diameter, or 2000 Å to 200 Å.

With a modern electron microscope using a double condenser lens system to illuminate the specimen the electron beam may be a few microns across. In one experimental system the x-rays generated in this way have been detected inside an electron microscope. This localized area is some 10 × worse than that already reached with the scanning microanalyser and must be used with thin specimens. So far it has not been possible to reduce the condensed illumination below one micron with the long working distance between the lens and object in the electron microscope.

An alternative scheme for illuminating a very small region of the specimen for analysis is currently under investigation at Cambridge by the author, and is shown in fig. 8. A normal electron microscope, shown with three lenses for convenience, is used to form an enlarged image of the specimen on the fluorescent screen, with a magnification of 5000 × followed by an optical telescope of 10 × outside the viewing window. For microanalysis an additional electron gun is placed below the fluorescent screen and is connected to the same high voltage source used for the normal imaging gun. The electron beam from this bottom gun is reduced by the same lenses that have magnified the original image, since the beam is travelling in the opposite direction. The rotation of the ray paths is not in the same direction, but this is of no consequence when imaging a round source into a round image. The depth of focus of the electron microscope is very great due to the small imaging angles employed. This means that when the specimen image is focused in the normal way on the final fluorescent screen, the electron source a few inches below this screen is in turn focused in the plane of the specimen. The amount of reduction is the same as the magnification of the original image and so a 100 micron source would be reduced 5000 × to 200 Å. A small hole in the centre of the fluorescent screen would allow this bottom beam to emerge. The x and y shift of the specimen stage would be used to move the chosen area of the image over this hole and then the analogous area of the specimen would be struck by the analysing electron beam. The emitted x-rays could be detected by a proportional counter in the normal way and analysed for the amount and nature of the elements present, with suitable corrections for the thinness of the specimen.

The astigmatism of the objective lens would be removed with a stigmator set in accordance with the fringes seen in the normal image. This same setting would hold for the bottom beam as well. Similarly, the aperture would be centred by viewing the final image and the upward beam would be monitored above the specimen to ensure that the most intense region passed through the objective aperture. If the two electron beams interacted or the x-rays from the broad area illuminated by the condenser entered the counter, then the top gun would be biased to cut-off during analysis and the bottom gun while viewing the image.

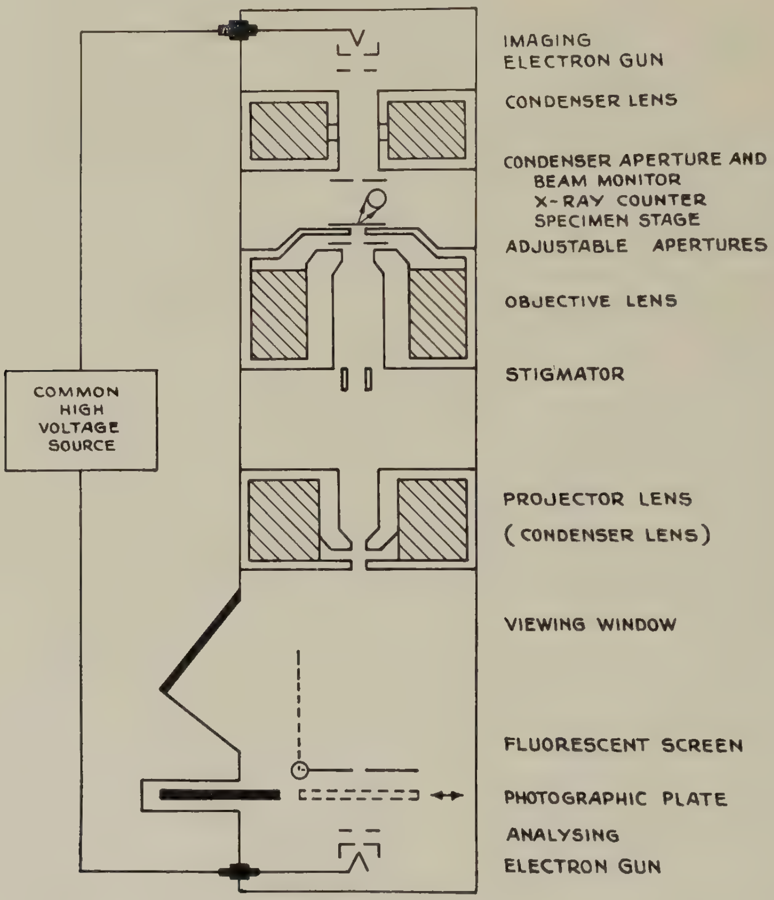


Fig. 8. A double gun electron microscope for selected area x-ray emission microanalysis at high resolution for transmission specimens.

The electrons striking the specimen will also be diffracted if the specimen is crystalline and produce radiation in the visible or ultraviolet regions as well as produce x-rays. The diffracted electrons could be imaged by the condenser lens onto a fluorescent screen and photographic plate in the plane of the anode of the imaging electron gun, giving a permanent record of the few diffracted spots from this very small selected area of the specimen.

In an ideal instrument it would be possible to analyse by emitted radiation for all elements from carbon upwards in the periodic table, with high accuracy and in low concentration, for an area of 100 Å diameter selected by viewing the already familiar electron microscope image of the chosen specimen. In this way quantitative electron microscopy would demonstrate microanalysis at the highest resolution and accomplish in one direction all that x-ray microscopy had set out to do.

7. CONCLUSIONS

Reflection x-ray microscopy has an affinity with optical mirror microscopy but seems to be the least promising method. Contact microradiography is simple and inexpensive and is widely used with a resolution approaching that of the optical microscope. Projection x-ray microscopy is more complicated but has an advantage where x-ray microanalysis is needed for quantitative results. Direct emission analysis with or without x-ray image formation is the most powerful of the methods described and may be expected to bring increasing accuracy into all areas of microscopical investigation at both the optical and electron optical levels.

REFERENCES

V. E. COSSLETT and W. C. NIXON, 1960, *X-ray Microscopy* (Cambridge University Press).

Proceedings of the Cambridge Conference on X-ray Microscopy and Microradiography, August 1956. 66 papers. (Academic Press, 1957.) Eds. V. E. COSSLETT, A. ENGSTRÖM and H. H. PATTEE.

Proceedings of the Stockholm Conference on X-ray Microscopy and X-ray Microanalysis, June 1959. 74 papers. (Elsevier, Amsterdam, 1960) Eds. A. ENGSTRÖM, H. H. PATTEE, and V. E. COSSLETT.

The Author:

W. C. Nixon received the M.A. degree in physics from Queen's University, Kingston, Canada, in 1949 and the Ph.D. degree from Cambridge University in 1952. He is an Assistant Director of Research in the Electronics Laboratory, Faculty of Engineering, Cambridge University and is engaged on various problems in applied electron physics.

The Friction of Solids

by E. H. FREITAG
Cavendish Laboratory, Cambridge*

On a misty morning when the wheels of a powerful modern diesel locomotive spin on the rails or when the car unexpectedly skids on a certain stretch of road we may suddenly learn to appreciate how much the friction between solid surfaces determines our daily activities. If it were not for such abrupt changes in conditions and experiences which are normally taken for granted only a few scientists and engineers would pause to think about the physical and chemical processes on which frictional behaviour depends.

The study of friction has at no time resulted in fierce controversies splitting the scientific world but it has nevertheless attracted great minds. Since the theories which are most widely accepted at present can be grasped without knowledge of their historical development a description of the pioneering work of Leonardo da Vinci, Amontons, Coulomb, Sir William Hardy, to name but a few, has been omitted from this article. Their achievements are implicit in the success of their disciples.

In reviewing the field it has been attempted to show that we are now able to understand, at least qualitatively, a great many aspects of frictional behaviour on the basis of a few simple mechanisms. Considering the staggering diversity of natural and artificially created surfaces that are at our disposal this may well be regarded as a surprising and remarkable result of scientific research. As a discussion of some problems of current and future investigations may suggest, the study of friction remains as exciting as that of more glamorous or more publicised branches of physics. Each small step forward opens a view into unknown territory where we can and must make full use of the methods and experimental techniques of modern physics and chemistry.

The basic ideas which serve to explain most phenomena associated with sliding and rolling are the fruit chiefly of research on metal surfaces. In order to understand the behaviour of other solids it is usually sufficient to modify a few of the most fundamental concepts.

The friction of metals

A friction force F can be measured whenever one body slides or rolls over another. The resistance to such motions is obviously connected with the fact that two surfaces are in contact, being pressed together by a load W. A study of the nature of the contact gives the first clues to where the source of friction lies and why sliding requires the expenditure of a considerable amount of work.

The Contact between Metals

It is important to realize that we are unable as yet to produce a perfectly flat metal surface. Even on the most carefully prepared specimens with a mirror-like finish deviations from the ideal plane can be detected which are large compared with the interatomic distance. The geometry of the smoothest surface can be examined by interferometric methods or by electron microscopy, but in many cases adequate information is more conveniently obtained by using a stylus instrument.

---

* Now at The Engineering Laboratory, Cambridge.

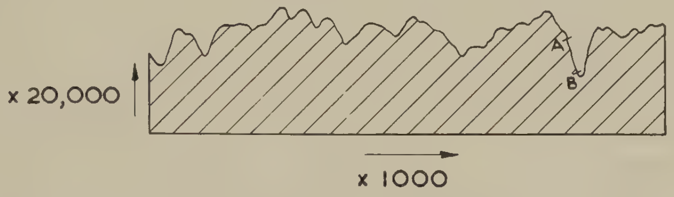


Fig. 1. Cross-section of finely ground steel plate. The profile was traced by a ' Talysurf ' stylus instrument. The true slope of portion AB is approx. 9°.

The cross-section of any finely ground, or polished metal surface normally looks like the profile drawn in (fig. 1). The contours suggest that such surfaces on a small scale, resemble a rugged landscape with randomly distributed hills and valleys. The inclination of the slopes, it should be remembered, rarely exceeds an angle of about 5°. All the same, the conclusion is inevitable that two metal surfaces cannot be perfectly matched.

As they are placed together they touch only at small localised regions. The load W has to be supported at precisely these areas, and this means that high pressures are developed locally even though the total normal force acting on the surfaces may also be quite small. It is to be expected, therefore, that the stresses in most regions where metal rests against metal soon exceed the elastic limit and cause plastic flow. This process obviously comes to an end when the area of intimate contact A created by this plastic deformation is large enough to support the load elastically.

The extent to which a pair of metal surfaces conforms in this manner depends, of course, on the applied load W and on the operating yield pressure p. To a first approximation the latter can be regarded as a material constant of about the same value as the indentation hardness of the softer of the two metals. Hence we can write

$$A=\frac{W}{p}$$

Measurements of the electrical resistance between two metal specimens have provided experimental evidence for the validity of this relationship (Bowden & Tabor 1954), and it is now generally believed that the true area of contact even for apparently flat and hard surfaces is created principally by the plastic deformation of irregularities. Thereafter the load is supported by the elastic stresses set up in the neighbourhood of the regions of intimate contact.

Thus the contact between two metals involves localised plastic flow as well as a reversible overall distortion of the surfaces. So far only stationary contacts have been considered but the arguments which apply when sliding or rolling takes place are substantially the same. In these cases a tangential force modifies and increases the contact stresses produced by the normal load so that A should be somewhat larger than when the surfaces are at rest.

In order to answer the question about the origin of this friction satisfactorily only one more basic concept needs to be introduced. A large body of experimental facts, collected mainly by Bowden and his colleagues at Cambridge

(Bowden and Tabor 1954) and subsequently confirmed by the work of many other scientists, suggests that the metals *adhere* at the regions of real contact. This ' cold-welding ' or formation of ' junctions ' occurs at room temperature and despite the presence of the oxide layers and films of adsorbed matter with which metals are covered in the normal atmosphere. To demonstrate the existence of such bonds one of a pair of surfaces has been made radioactive, and by measuring an emission from the other surface after sliding it could be shown that metal particles are exchanged during such a process and left adhering to the specimen opposite. Occasionally the transferred fragments are large enough to be sectioned by metallographic techniques and photographed with the metal to which they have become attached.

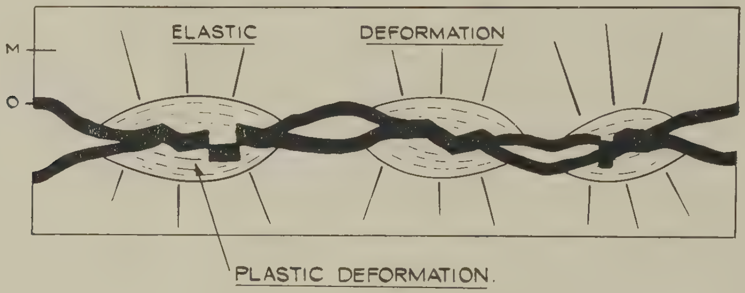


Fig. 2. Schematic view of the contact between metal surfaces in air. M = metal, O = oxide layer. The enclosed broken-shaded regions indicate plastic deformation and the vertical rays elastic deformation.

A schematic illustration of the contact between metal surfaces is shown in (fig. 2) and if adhesion is visualised it becomes obvious that relative motion can take place only when these bonds are continually broken. This is the proposition of the adhesion theory of friction.

The Adhesion Theory

With this simple picture of the contact conditions before us we readily see why sliding normally requires a much larger force than rolling. In sliding all the junctions must be sheared simultaneously whereas in rolling the surfaces are rather peeled apart. Thus the friction should be quite different in the two cases if adhesion is an important factor. In practice the sliding resistance for hard surfaces is often 1000 times greater than the rolling resistance.

There is some justification, therefore, in assuming that adhesion is the *chief* source of sliding friction. This supposition and the fact that the area of true contact is determined by the load W and the yield pressure p enable us to explain satisfactorily many aspects of frictional behaviour.

Two empirical laws concerning the variation of sliding friction (usually named after Amontons) hold over a remarkably wide range of conditions :

(1) The friction force F is proportional to the load W.

(2) F is independent of the size of the apparent contact area.

Introducing s as the average shear strength of the junctions and regarding it as a constant we can express our assumption about the origin of the friction force by

$$F = sA$$

and since

$$A = W/p$$

we get

$$F = \frac{s}{p} W,$$

or defining the coefficient of friction μ in the usual way,

$$\mu = \frac{F}{W} = \frac{s}{p}.$$

Hence the coefficient of friction is a material constant as Amontons' laws imply.

Plastic Deformation

This simple view of the friction mechanism seems adequate in the sense that the most generally observed rules about the sliding behaviour of metals can be derived from it. It also helps us to understand why the friction does not markedly change when the speed is varied or when the metals are uniformly heated to a higher temperature. For both the shear strength s and the yield pressure p are properties connected with plastic deformation and therefore affected in the same way when the conditions are changed. (If extreme sliding speeds or temperatures are imposed this remains no longer true, of course, since new factors such as frictional heating and diffusion can play an important part and alter the sliding mechanism.) The same considerations apply also to a comparison of the friction for hard and for soft surfaces. The ratio s/p can be the same in both cases.

The limitations of this elementary treatment of the friction process become at once apparent when we try to predict the value of μ for a given pair of surfaces. It is impossible to calculate the relative magnitude of s and p since the effects of the normal and tangential forces on the deformation of an individual junction cannot be separated. The metal yields when the *combined* stresses exceed the elastic limit. The process of shearing is further complicated by the fact that the interfacial region is not purely metallic but contains weakening constituents such as oxides and adsorbed matter. Thus each junction is deformed plastically in a manner not accurately known at present, and it breaks when the interface fails to transmit a certain shear stress. Clearly, the processes are basically the same for clean and contaminated surfaces, and the main difference lies in the earlier failure of the contaminated contact regions. A detailed discussion of the deformation mechanism has recently been published by Tabor (1959). Despite these real difficulties the adhesion theory provides reasonable explanation for most observations on the sliding behaviour of metals. That strong junctions are formed has been suggested by the evidence of transfer. To emphasize their importance as compared with other possible sources of friction seems justified if we consider, for instance, the profound effect which minute quantities of organic substances have on the coefficient of friction.

For pure metal surfaces, as they may be obtained by heating *in vacuo* until evaporation sets in, the value of μ can lie between 10 and 100. Theoretically, sliding should be impossible ($\mu = \infty$) under these conditions since the interface may be as strong as the bulk metal so that two ductile specimens should be 'welded' into one piece by the combined normal and tangential stresses. In the normal atmosphere the junction between clean metals contains the weakening oxides and adsorbed molecules which are inevitably present on all surfaces. Hence the coefficient of friction falls into the range of about 0·3 to 1·5. A further reduction by a factor of 10 can be achieved simply by covering the surfaces with suitable organic films. A single layer of molecules (of a fatty acid for instance) is sufficient to produce this lubricating effect.

Such a treatment can hardly alter the topography or the plastic properties of the metals to a significant extent but it must have a considerable influence on the interaction of the surface forces which determine the adhesion. Long molecules provide a better shield against mutual attraction than short ones of the same type of compound, and as soon as the 'boundary' film has been removed either mechanically or by heating, the friction increases again to its normal value.

These observations, and many which have not been mentioned, leave little doubt that in sliding adhesion is the main cause of friction between comparatively smooth surfaces of similar hardness. Where these conditions do not exist other factors have to be taken into account. Since the surfaces are never perfectly flat the irregularities will necessarily impede each other, and in the case of a large difference in hardness between the two metals the hard asperities will plough or even cut through the soft material. However, these processes also involve much plastic flow. It will be difficult in general to separate the work expended in this type of deformation or abrasive action from that required for shearing junctions.

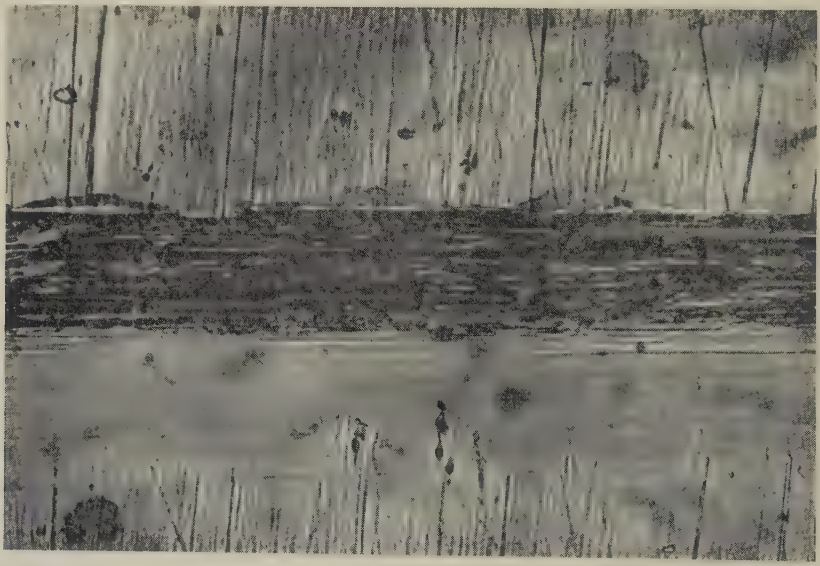


Fig. 3. Friction track on copper surface, produced by a hemispherical steel slider which traced the wide horizontal groove. The effects of ploughing and of tearing due to adhesion (dark areas) are both visible.

Whatever the detailed mechanism it seems unavoidable that sliding causes some damage to the surfaces. Unless the load is extremely light many small topographical features will suffer a permanent distortion, and where adhesion takes place heavy tearing can occur as the interface may be stronger than the metal on either side. Strong bonds of this kind are formed particularly between similar metals. In fig. 3 the effects of both ploughing and tearing are illustrated. The groove was produced by a hemispherical steel rider on copper, and the heavy damage in certain areas is due to copper which had been picked up by the upper surface during the earlier stages of the experiment. Quantitative wear studies by Archard (1958) have shown that in general the majority of junctions break at the interface.

Elastic Deformation

Although the indirect evidence about the rôle of adhesion in sliding seems quite convincing it has often been critically asked why the direct measurement of the effect frequently presents great difficulties. If two surfaces readily stick together in the ordinary atmosphere it should be possible to determine the strength of the bond simply by removing the load and then applying a negative normal force to separate them. The result of such tests is usually disappointing and probably for the following reason. As the contact pressure is reduced the surfaces alter their shape by elastic recovery. The relative movements caused by this release of strain may be of only microscopic dimensions but few of the small work-hardened junctions can be stretched to the extent required. Thus they may be broken individually so that the surfaces peel apart even before a negative load has been applied.

It has been pointed out earlier that a similar sequence of events takes place in rolling. In this type of friction process the influence of junctions has nearly disappeared because no appreciable slip occurs within the area of contact. As the load traverses the lower surface there is compression ahead of the rider and recovery behind it. This elastic deformation is accompanied by a small loss of energy which would show up as a hysteresis loop on a stress-strain diagram. Consequently, even the reversible changes give rise to a friction force. If the load is not high enough to cause the formation of a permanent groove, elastic hysteresis constitutes the chief rolling resistance. The friction in these cases is so small that in a ball-bearing (which is designed to exploit this feature) the energy loss is mainly due to some inevitable rubbing between the balls and the cage or, particularly at high speeds, to the shearing of the lubricant (Palmgren 1936).

In the sliding of metals this effect contributes only a negligible proportion of the total friction force but with viscoelastic solids it can overshadow all other sources of friction provided the interfacial adhesion has been reduced by lubrication. Some practical implications of elastic hysteresis in skidding friction of tyres on roads have been described in detail by Miss Sabey (1959).

These views on the relative importance of adhesion, abrasion and elastic hysteresis for the friction of metals can undoubtedly be challenged in various ways. On the basis of the present state of knowledge it may fairly be claimed, however, that the phenomena associated with sliding are most satisfactorily explained by the adhesion theory.

THE FRICTION OF NON-METALS

The discussion of the friction mechanism for metals would appear to be quite generally valid as far as it suggests that the behaviour of all solids in sliding and rolling processes depends upon their deformation and surface properties.

Considering the diversity of these properties even in a narrow selection of non-metals (comprising, for instance, inorganic brittle crystals, polymers and wood) the formulation of a general theory does not promise to be a rewarding undertaking. Nevertheless the principal elements of the adhesion theory have been successfully applied to all these types of solids. The modifications which are necessary for a given pair of surfaces concern in most cases the relationship between the load and the area of true contact.

Surface roughness must again be regarded as a general characteristic. Molecularly flat areas of appreciable size have been observed only on cleaved mica sheets (fig. 4). With many solids (rubber and diamond are obvious examples) the deformation of the irregularities is elastic rather than plastic, or it may follow an intermediate law. But even in these cases A grows nearly proportionally with W provided the number of contacts is large and increasing with the load, (Archard 1957).

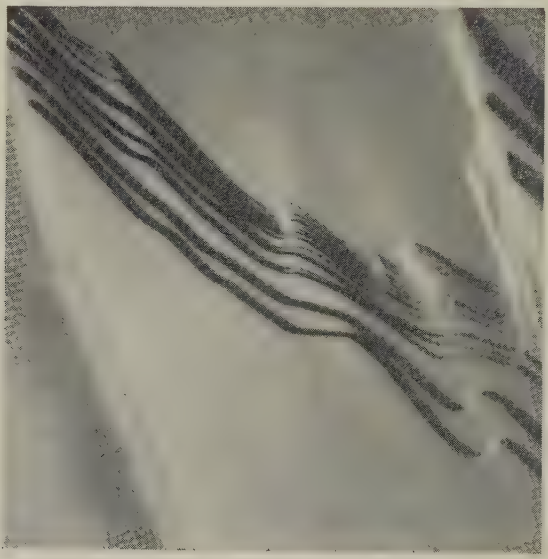


Fig. 4. Electron reflection micrograph of mica surface. The height of the steps is a multiple of the spacing of the molecular lattice. Horizontal magnification 300 ×, vertical magnification 3000 ×.

Where the non-metallic surfaces are pressed into close proximity the molecular fields of force overlap so that a strong mutual attraction results. Since these bonds have to be sheared during sliding adhesion plays again an important rôle in the friction process. At contacts between two pieces of the same material the interfacial forces will resemble the cohesion within the solid, and the separation of such junctions may be expected to occur occasionally in the bulk rather than at the interface. This naturally leads to the transfer of detached particles.

Investigating a group of linear polymers such as polythene, polyvinylchloride etc., Shooter and Tabor (1952) found indeed that the friction force is approximately equal to the product of the area of contact and the shear strength of the plastic in bulk. That the interface is often stronger than the softer of the two materials could be demonstrated even with polymers sliding on metals and, what is more remarkable still, a plastic such as nylon is sometimes capable of plucking out a small fragment from a hard steel surface (Rabinowicz and Shooter, 1952), thus causing some wear.

Polytetrafluoroethylene is noted for its intrinsically weak adhesion to other solids and for its low coefficient of friction. μ lies usually around 0·1 while for other similiar polymers values between 0·2 and 0·5 are measured. An explanation of these properties in terms of the molecular structure of PTFE has been put forward by Bunn and Howells (1954) but in general very little is known as yet about such correlations.

Many polymers show markedly visco-static behaviour and are rather easily deformed so that hysteresis losses contribute a much larger fraction of the total sliding resistance than in the case of metals.

The amount of energy which is dissipated internally depends, of course, on the rate of deformation. Hence the friction can be markedly sensitive to speed. For lubricated sliding and for rolling where the effect of adhesive bonds is greatly minimized some success has been achieved in relating the hysteresis losses, and thereby the coefficient of friction, to the mechanical bulk properties of the materials (Bueche and Flom 1959 ; Greenwood, Minshall and Tabor 1960).

In analyzing the results of friction experiments with non-metallic solids it is therefore often useful to express the tangential force F as the sum of the two independent terms

$$F = F_H + F_A \, ,$$

where F_H represents the force due to deformation with hysteresis loss, and F_A denotes the force required to overcome interfacial adhesion.

The work of Atack and Tabor (1958) on balsam wood suggests that both components are affected by the moisture content but that our simple model of the friction mechanism is valid even for materials with such a complex structure and composition.

The study of brittle crystalline solids raises another issue. One might presume that the adhesion theory as it has been presented so far does not apply to these materials since it involves either plastic flow (with metals) or visco-elastic deformation. Yet despite some cracking a friction track in rock-salt gives the impression of predominantly plastic behaviour, and the shear strength of the junctions, as calculated by King and Tabor (1954), is several times greater than that of a single crystal. As many crystals acquire considerable ductility if subjected to hydrostatic compression it is concluded that the compressive stresses around each individual contact inhibit its brittle disintegration during sliding. Hence, unless the materials are excessively brittle, their low shear fracture strength does not substantially affect the friction mechanism.

MODERN RESEARCH DEVELOPMENTS

In the first part of this article the friction of solids has been discussed in such elementary terms that a great many questions about various details of the sliding mechanism may have cropped up in the reader's mind. Some of these aspects have recently been studied or are being investigated but many problems remain to be tackled in future research work.

A more critical approach to the subject makes one realize that even the pillars of the adhesion theory are not as reliable as one could wish. There exists, for instance, no really satisfactory method for measuring the true area of contact accurately. While it must be admitted that to fill this gap would tax the ingenuity of the best experimenters it would be valuable to know to what extent the interface between hard well-matched surfaces is determined by elastic deformation.

Adhesion has been recognized as the principal source of sliding friction, yet our knowledge of its mechanism is still very limited. Contamination or boundary lubricant films weaken the bonds considerably but while the friction is comparatively insensitive to such changes and may drop to a tenth of its former value the amount of transfer and wear can be reduced tenthousandfold or more. From this point of view the main issue is whether the contacts are broken by separation at the interface or by shearing in the bulk. Hence a small difference in the strength of junctions can obviously have far-reaching consequences.

On many of the basic aspects of adhesion few experiments have been carried out up till now, and the major ideas have been deduced from friction and wear studies. Some of the modern Russian research is still strongly oriented in this direction (Kuznetsov 1960). In several laboratories efforts are now being made to obtain fundamental information by more direct methods. Working with large single crystals of copper which were placed together in a hydrogen atmosphere and heated to about 800°c Dyer and Gwathmey (1959) found that the relative crystallographic orientation of the specimens was an important factor in the bonding process. Similarly, according to Semenoff (1958) strong adhesion requires a favourable mutual arrangement of the two lattices.

When metals are pressed together in air the oxide layers may be disrupted during the plastic deformation, and the extent of the breakdown will be influenced by the relative hardness and the relative ductility of the oxide and the underlying metal. It has often been observed that interfacial slip which might assist in rupturing the protective film clearly helps the formation of strong bonds, but whether purely metallic contact in a few places is essential for appreciable adhesion cannot be stated with certainty at present.

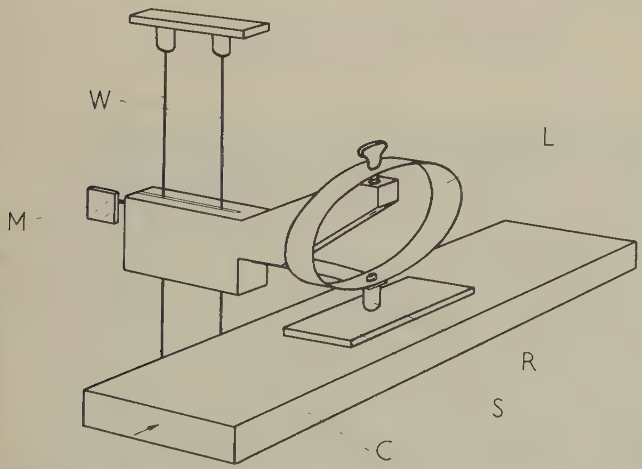
It is a common experience for scientists in this field to get disappointed by the poor reproducibility of the results which is indicative, of course, of the lack of understanding of the controlling variables. However, by using statistical methods to interpret a series of simple tests Anderson (1959) was able to verify that friction and adhesion are interrelated phenomena. In these experiments the butt-ends of two cylinders were pushed together and then, after the rods had been twisted under load, separated by coaxial tension. The results show that if the adhesive force is substituted for the tangential force in Amontons' laws the same rules hold for such bonding processes. It is proposed to study the effect of surface roughness, hardness, elapsed time, amount of interfacial slip by the same technique.

Apart from this urgently needed information the solution of a number of fundamental physical problems is eagerly awaited. Experimental data on the range of surface forces and on the relation between chemical structure and adhesion could throw much light on the friction mechanism. The rôle of the oxide and its structure on the bonding of metals could also be the subject of a useful inquiry. Considerable technical difficulties may have prevented a thorough investigation of electrostatic effects in connexion with friction and adhesion but in view of their practical significance, in the processing of synthetic fibres for instance, this surely cannot be delayed much longer.

In applied as well as in pure science there is a tendency to-day to explore the extremes of the range of conditions which can be imposed on a process, and the recent studies of friction at very high speeds, and at very high temperatures may serve as examples.

The Friction at Very High Speeds.

The work expended in the sliding process is dissipated mainly as heat. As the junctions are sheared the temperature of the deformed metal can rise to surprisingly high values. The heating is, of course, confined to small localized regions so that temperatures exceeding several hundred degrees c can be produced although only a moderate load of about 1000 g and a sliding speed no greater than 100 cm/sec have been applied. Since non-metallic solids are poor conductors of heat this effect occurs even more readily. By rubbing a metal pin against a glass disk bright luminous ‘ hot spots ’ can be made visible within the area of contact. This raises the question of scientific as well as practical interest : Could surfaces be melted by using a very high sliding speed, and what kind of friction mechanism would operate under these conditions ?



C=Movable Support
L=Loading Spring
M=Mirror
R=Rider
S=Specimen
W=Bifilar Suspension

Fig. 5. Diagram of a conventional friction apparatus (designed by Bowden and Leben).

Friction measurements are usually made with very simple equipment. A means for pressing two surfaces together and a sensitive device for measuring the tangential force during sliding are the basic requirements. Fig. 5 shows a diagram of a type of apparatus which has been used in many pioneering researches.

For experiments at very high speeds another approach is obviously needed. Surface speeds in the range of ballistic velocities can be produced conveniently by a method developed by Beams (1942) for spinning ultracentrifuges. A steel rotor is suspended and accelerated by magnetic fields *in vacuo* so that the torsional resistance is very small. At present surface speeds of about 10^4 cm/sec are obtainable in this way, the limit being determined by the strength of the rotor.

In our friction studies (Bowden and Freitag 1958) a commercial steel ball was used and placed between three symetrically arranged vertical surfaces. These flat specimens were mounted in such positions that they could not interfere with the rotor during its acceleration. The level at which the spinning ball floated could be controlled by a photo-electric feed-back system as indicated n fig. 6 and a stabiliser was built into the friction device to prevent precession and lateral vibrations. The surface of the ball was marked so that the intensity of the light reflected from it varied sharply once every revolution. These signals were then picked up by a photo-multiplier cell and counted electronically. By feeding the result into a recorder a plot of the rotational speed against time was obtained. As soon as the ball had reached the desired speed a spring was released and the ball trapped as shown in fig. 6. Thus the same load was applied to all three contacts. Friction then caused the rotor to slow down rapidly and since the rate of deceleration was known from the speed-time record it could easily be calculated. In order to preserve the high-speed wear marks some experiments were interrupted by withdrawing the spring-operated surface before the ball had lost much energy.

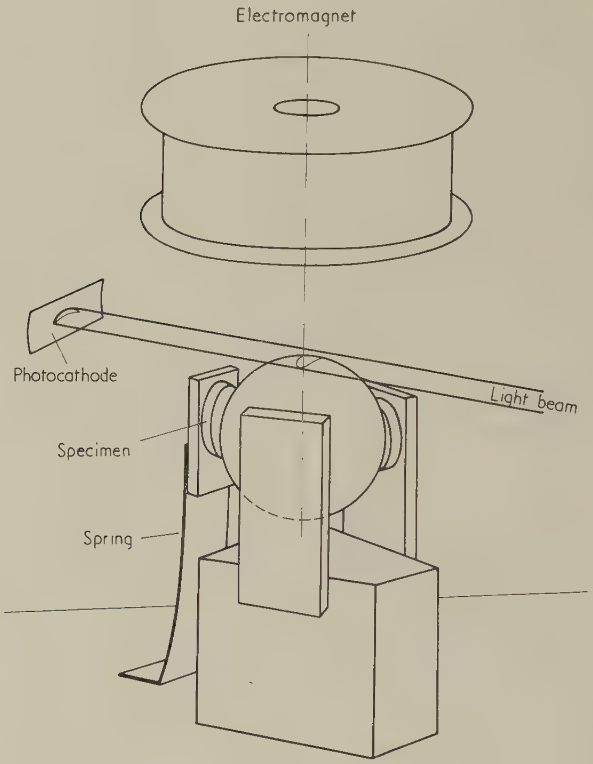


Fig 6. Arrangement for measuring friction at very high speeds. The steel ball is suspended electromagnetically and accelerated by a rotating magnetic field.

Examining the surface damage we discover at once that the result of very rapid sliding is quite different from the usual experience. A steel ball rubbing against copper at 300 m/sec. produces plastic flow on a large scale in contrast to the abrasion and tearing which are characteristic for ordinary conditions (compare figs. 7 and 3). At 600 m/sec. the apparent area of contact receives a smooth shiny polish which suggests that the frictional heat has melted a thin layer of metal (fig. 8). By inserting fine thermocouple wires into the contact area it has been possible to measure flash temperatures as high as the melting point of copper. An analysis of the heat flow in the specimens shows, however,

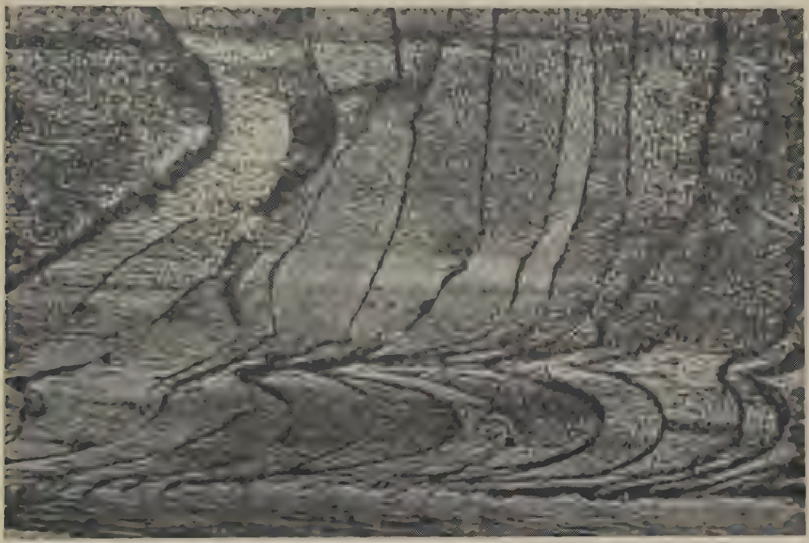


Fig. 7. High-speed wear mark on copper. Sliding at 200 m/sec caused plastic flow on large scale. Magnification × 150.

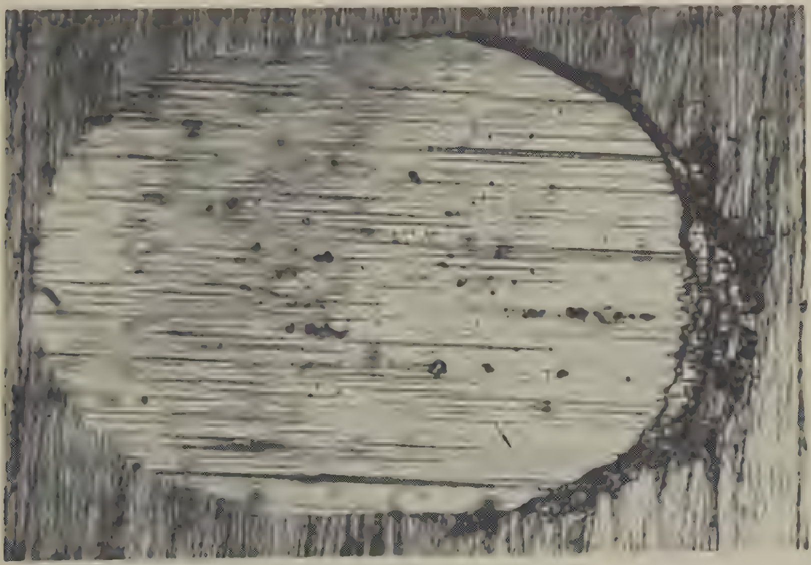


Fig 8. High-speed wear mark on copper. Sliding at 600 m/sec produced surface flow and a smooth finish. Magnification × 250.

that only very localised melting could have occurred in this case. At a load of about 50 g the spinning ball does not supply enough energy to keep a layer of the size of the apparent area of contact molten. It must be concluded, therefore, that only small regions are intensely heated by the rapid deformation. Since the heat is conducted away at a high rate from the junctions a steep temperature gradient is set up near the interface. Hence the shear strength of the junctions is reduced by a greater amount than their load-carrying capacity. Consequently, an increase in the sliding speed leads to a lower coefficient of friction, and this is borne out by the results plotted in fig. 9. All these experiments have been carried out *in vacuo* where ductile metals normally seize together because the rubbing action tends to remove the oxide and contaminant films. The measurements were made within a second after establishing contact as prolonged sliding at a high speed can change the coefficient of friction.

According to fig. 9 bismuth seems to behave like copper although the friction is much lower. The full range of available sliding speeds could not be used, however, since a shower of fragments and drops of metal was ejected from the contact area when the speed exceeded about 300 m/sec. Thus the beam of light needed for measuring friction was cut off by the flying debris. It was clear though, and fig. 10 provides convincing evidence, that metals with a low melting point and poor conductivity can be melted by rapid sliding. The technique just described is obviously inadequate for studying the mechanism under these conditions.

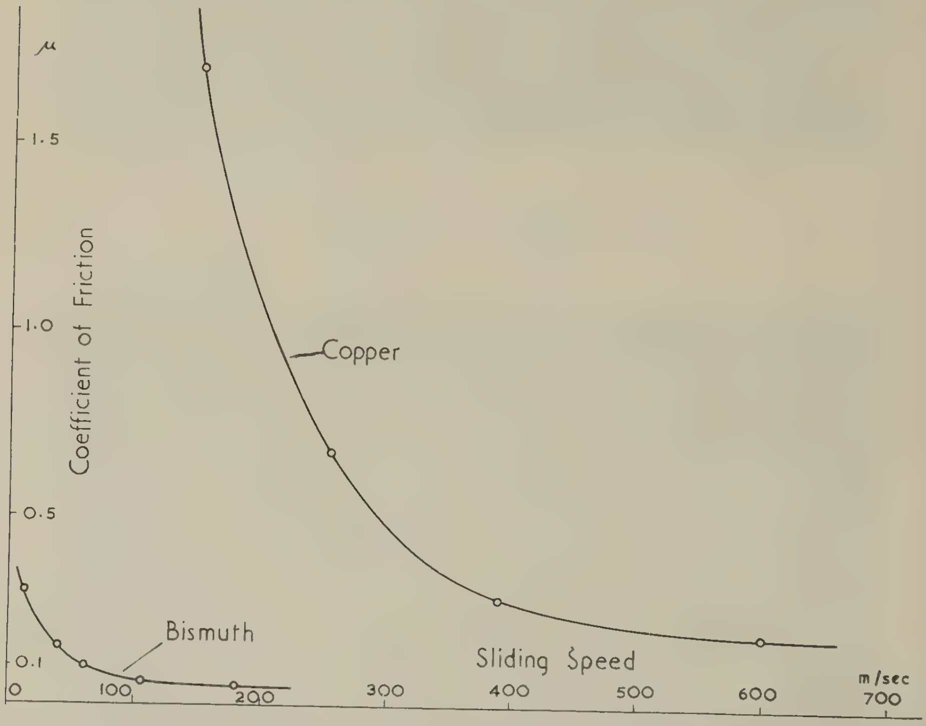


Fig. 9. The coefficient of friction for steel sliding on copper and bismuth (*in vacuo*). Loads 30 g and 10 g, respectively.

It has been replaced (Bowden and Persson 1960) by the simple method of dropping a spinning ball on to an inclined flat plate as the diagram in fig. 11 indicates. Without rotational energy the ball would rebound in the normal plane (O C A) but because of the friction force experienced during the impact on the specimen it is deflected sideways by an angle α. If the ball is stopped by a screen a second mark (B) is produced and α can be obtained simply by measuring the distance l from B to the intersection with the plane (O C A). The sliding speed depends, of course, on the rotational speed of the ball and on the inclination of the target. It can be regarded as constant since it changes only by a fraction of a per cent during the short time of contact t (about 1 m/sec.), As both the friction force F and the load W vary considerably during the impact, both generally in a similar manner, it seems reasonable to define

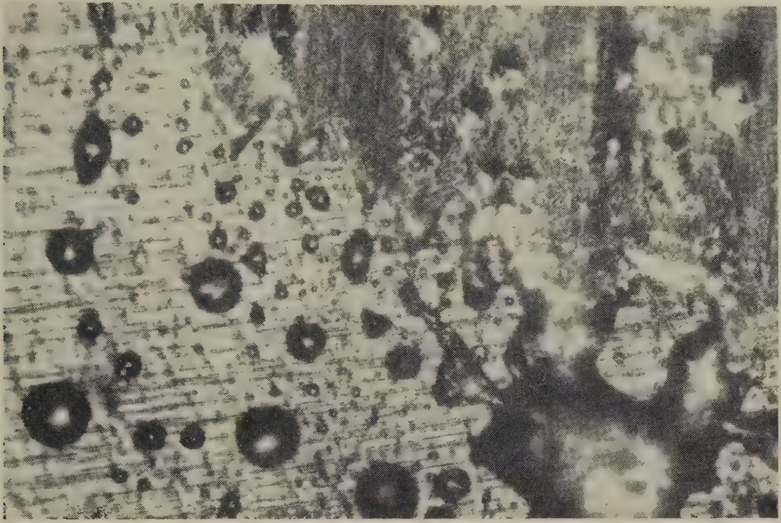


Fig. 10. High speed wear mark on bismuth. The metal was melted by friction. Sliding speed approx. 350 m/sec. Magnification × 250.

$$\bar{\mu} = \frac{\int_0^t F}{\int_0^t W}$$

as the average coefficient of friction for these sliding conditions. Simple calculations show that μ, to a good approximation, is directly proportional to the deflection α.

For those speeds where the results of the two kinds of experiment can be compared the coefficients of friction agree remarkably well although the load rises to a peak of over 1000 g during the impact. From the second investigation we learn, however, that the friction increases sharply when the ball rides on a liquid film. Fig. 12 shows this tendency for steel sliding on bismuth and Wood's alloy. Due to the large-scale melting the surfaces are worn very rapidly in this speed range.

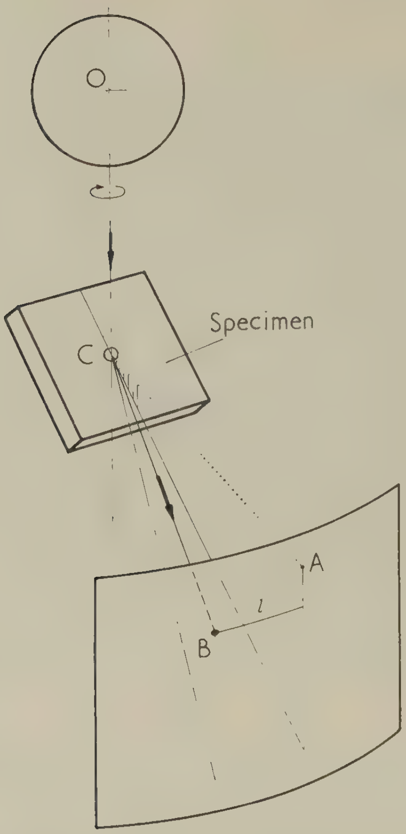


Fig. 11. Diagram illustrating the technique of measuring friction at very high speeds by dropping a rotating sphere.

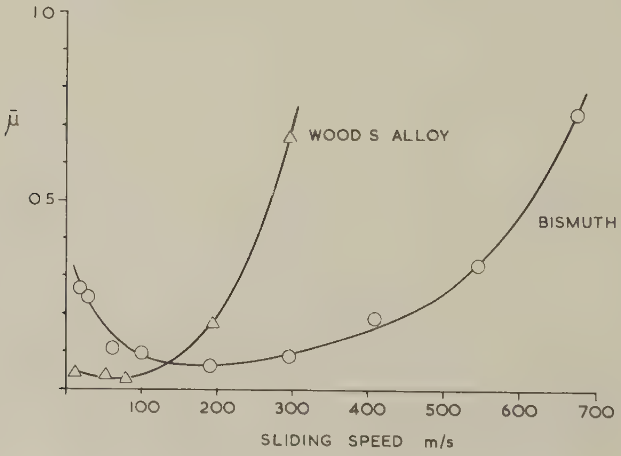


Fig. 12. The coefficient of friction for steel sliding on bismuth and Wood's alloy, respectively. Results of impact experiments according to fig. 11.

The mechanism which emerges from these observations and results can be stated in almost trivial terms. The whole apparent area of contact can be melted by rapid sliding provided enough frictional heat is generated to

(1) supply the energy for melting more metal. This is necessary in order to replace the liquid which is continually wiped away from the interface.
(2) balance the energy lost mainly by conduction.

For metals the experimental evidence agreed well with the predictions based on approximate energy calculations while the behaviour of other solids does not fit into this simple theory. The coefficient of friction of nylon remains low even when the interface is undoubtedly molten. Since the viscosity at the melting temperature is very high in this case much energy could be absorbed in shearing the liquid film. This would reduce the viscosity and thus tend to keep the friction low despite increasing speed. This effect clearly merits further attention.

Friction at Very High Temperatures.

Operating temperatures in power producing plants and in propulsion aggregates have been increased steadily in the past and the demands for advances in this direction are pressed with growing vigour. Accordingly new heat resistant materials with many unknown properties are continuously tried out in the bearings and seals of moving parts. Their friction and wear characteristics are thus no longer only of academic interest but assume considerable practical importance.

Friction studies at very high temperatures are carried out in specially designed furnaces. Figure 13 shows a diagram of the equipment built by Rowe (1958) and

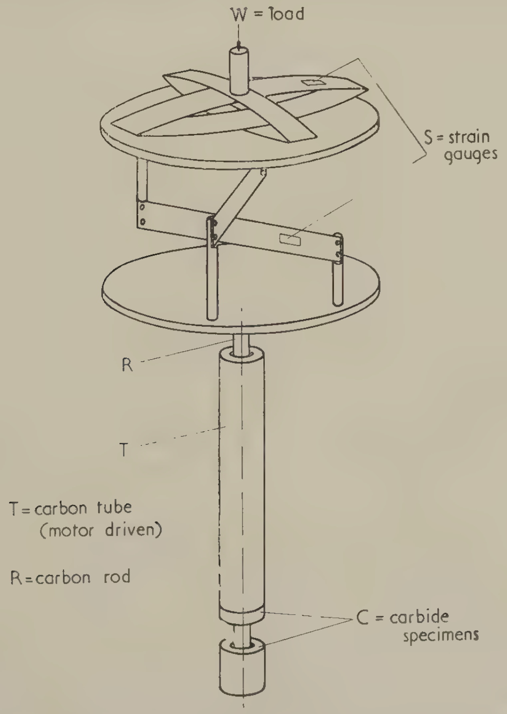


Fig. 13. Schematic view of apparatus for measuring friction at very high temperatures (design Rowe).

later used by Mordike (1960). In this type of apparatus two anular specimens can be rotated relative to one another at temperatures up to 2000°c *in vacuo*. The load and the torque are applied by a system of crossed springs and measured by strain gauges. In Mordike's experiments the sliding between nominally flat surfaces of sintered carbide and boride specimens was investigated, and this was followed by some work on the friction between graphite and a series of refractory materials. The results plotted in fig. 14 are typical for the behaviour of a pair of carbide surfaces. As the temperature is raised the coefficient of friction decreases gradually until, at about 1000°c, the trend is abruptly reversed. In the region of 2000°c the value of μ is 2 or 3 times higher than at room temperature. This variation can be accounted for if it is supposed that the real area of contact is initially determined by elastic deformation. Since we are dealing with extremely hard and brittle materials this view seems preferable in this case to the general mechanism. Hence a rise in temperature should not alter the size of the junctions but only weaken the interfacial adhesion. This would explain the fact that the coefficient of friction is reduced at first. At about 1000°c the surfaces begin to sinter together and much stronger bonds are formed because diffusion and increased plasticity of the crystallites play a major part at very high temperatures.

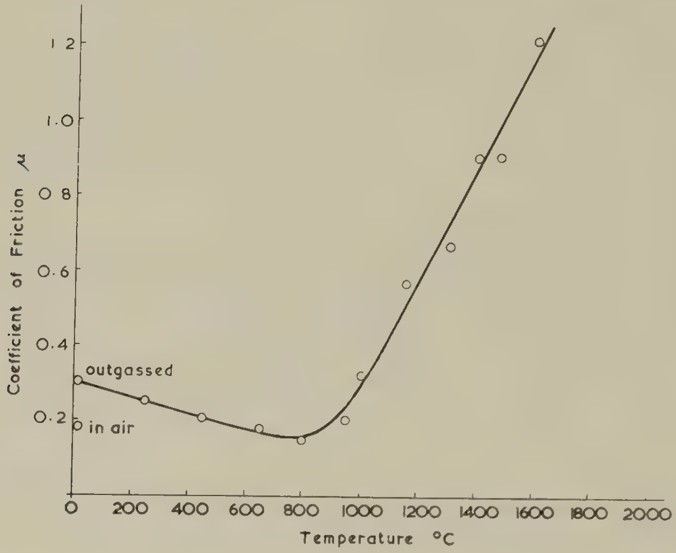


Fig. 14. The coefficient of friction for the sliding of niobium carbide on vanadium carbide as a function of temperature. Load 1500 g.

According to the graph fig. 15 the friction does not change in the described manner when a carbide specimen slides on graphite. Compared with most refractory materials graphite possesses a very low shear fracture strength, and this may be the dominant factor over the whole temperature range. When graphite rubs against graphite the friction steadily decreases on heating.

A similar investigation of the sliding behaviour of hard heat resistant compounds has been reported recently by Zeman and Coffin (1960). They adopted a simple sliding system of crossed cylinders and limited the temperature to about

1000°c so that the friction of these solids could be studied in air. The materials were selected from a very wide range since it was hoped that the results might reveal a correlation between the sliding characteristics and the solid solubility or other basic physical or chemical properties. For metals Roach, Goodzeit and Hunnicutt (1956) had found that the propensity of a given pair of surfaces to seize and score under severe sliding conditions where high local temperatures are developed can be predicted from the relative position of the metals in the periodic table. So far it has not been possible to establish a fundamental criterion for the behaviour of refractory compounds.

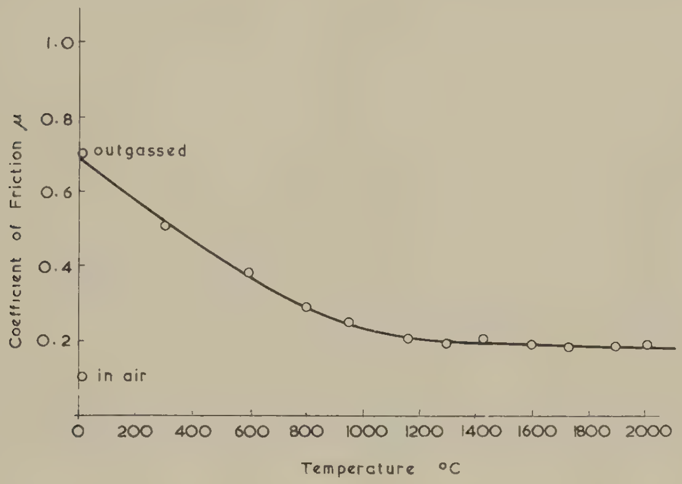


Fig. 15. The coefficient of friction for the sliding of graphite on niobium carbide as a function of temperature. Load 1500 g.

An interesting result of this work is the fact that certain oxidation products formed at elevated temperatures, particularly B_2O_3 on boron carbide, can act as lubricants. However, when the surfaces were cooled afterwards, this beneficial effect did not persist. With other hard solids continuous smooth sliding was prevented by abrasive debris which had accumulated.

The friction properties of these heat resistant materials are obviously not yet fully understood and future progress will depend much on how rapidly new data on their strength properties at high temperatures become available.

In conclusion it might again be pointed out that the general trends of frictional behaviour can now be understood on the basis of a simple scientific theory. The applications of this theory, while not promising such revolutionary changes as nuclear research and the exploration of space are believed to bring about, have influenced modern engineering practice to a remarkable extent. All the same, our knowledge of the interaction between surface asperities and their subsequent deformation is still so incomplete that scientists can sometimes give but little help with practical difficulties. The need to fill these gaps is great indeed as many technological developments, in the field of rockets and missiles for instance, are retarded not so much by major physical problems but by inexplicable friction and wear phenomena. Indirectly, friction and adhesion studies have often yielded useful information on various bulk and surface properties of solids, and it is to be hoped that progress in this field will continue at the pace set in other branches of science.

REFERENCES

ANDERSON, O. L., *J. App. Phys.* **30** (1959), 593.
ARCHARD, J. F., *J. App. Phys.* **24** (1953), 981.
ARCHARD, J. F., *Proc. Roy. Soc.* A. **243** (1958) 190
ATACK, D. and D. TABOR, *Proc. Roy. Soc.* A **246** (1958), 539
BEAMS, J. W., J. WASH. *Acad. Sci.* **37** (1947), 221.
BOWDEN, F. P. and D. TABOR, *The Friction and Lubrication of Solids*, 2nd ed. (1954), O.U.P.
BOWDEN, F. P. and E. H. FREITAG, *Proc. Roy. Soc.* A **248** (1958), 350.
BOWDEN, F. P. and P. A. PERSSON, *Proc. Roy. Soc.* A **261** (1961), *in print.*
BUECHE, A. M. and D. G. FLOM, *Wear* **2** (1959), 168.
BUNN, C. W. and E. R. HOWELLS, *Nature* **174** (1954) 549.
GREENWOOD, J. A., H. MINSHALL and D. TABOR, *Proc. Roy. Soc.* A **259** (1961), 480.
KING, R. F. and D. TABOR, *Proc. Roy. Soc.* A **223** (1954), 225.
KUZNETSOV, V. D., *Metal Transfer and Build-up in Friction and Cutting* (1961), Pergamon Institute, *in print.* (Russian ed. Moscow 1956.)
MORDIKE, B. L., *Wear* **3**, No. 5 (1960), 374.
PALMGREN, A., *Ball and Roller Bearing Engineering* (1945), *Philadelphia.*
RABINOWICZ, E. and K. V. SHOOTER, *Proc. Phys. Soc.* B **65** (1952), 671.
ROACH, E. A., C. L. GOODZEIT and R. P. HUNNICUTT, *Trans. A.S.M.E.* **78** (1956), 1659.
ROWE, G. W. and D. J. BALDWIN, T.I. *Res. Labs. Report* No. **62** (1958).
SABEY, B., *Contemporary Physics*, **1** (1959) 56.
SEMENOV, A. P., "*Sativanye Metallov*", Mashgiz, Moscow (1958).
SHOOTER, K. V. and D. TABOR, *Proc. Phys. Soc.* **65** (1952) 661.
TABOR, D., *Proc. Roy. Soc.* A **251** (1959) carbide 378.
ZEMAN, K. P. and L. F. COFFIN, G. E. *Res. Lab. Report* No. 60-RL-2347 M (1960).

The Author:

Dr. E. H. Freitag graduated with the Diploma in Mechanical Engineering at the E.T.H., Zurich, in 1950. He worked for several years in the Department of Physics and Chemistry of Solids at the Cavendish Laboratory. Since October 1959 he has been Assistant Director of Research at the Engineering Laboratory, Cambridge. He is interested chiefly in the properties of engineering materials.

The Physics of Diamond

by Dr. E. W. J. MITCHELL
University of Reading

A two day conference on The Physics of Diamond was held on September 22nd and 23rd, 1960 at the J. J. Thomson Physical Laboratory of the University of Reading. The Conference was sponsored by the Institute of Physics and Physical Society and about 120 people attended from eight countries.

Professor R. W. Ditchburn in his introductory remarks pointed out how much diamond research, particularly in Britain, was helped by the co-operation of the Diamond Research Laboratory, Johannesburg, in providing specimens to meet individual requirements. There followed a paper by Dr. E. Lundblad (Sweden) describing the successful experiments at A.S.E.A., Stockholm on the synthesis of diamond. These were first carried out in 1953 and involved the precipitation of carbon from iron/carbon solutions, or from iron carbide (Fe_3C), at pressures and temperatures in the region of 10^5 atm and 4000°c. Dr. Lundblad reported that they had not been successful in obtaining the direct graphite/diamond transition.

During the discussion of this paper Professor Kennedy (U.S.A.) referred to some recent measurements of the absolute pressures of the phase transition (e.g. bismuth) used in most high pressure research as pressure indicators. He mentioned some conflicting points in Bridgman's calibrations depending on whether resistance or volume changes were used to follow the phase transition. The piston in Kennedy's experiments could be rotated at pressure in order to relieve the frictional effects which he claimed were the source of the inconsistencies in Bridgman's calibration. The measurements reported by Kennedy indicate that true pressures above about 20 000 atm are 25–30% lower than those based on the pressure scale hitherto used. Attention was also drawn to the fact that changes in the pressure calibration would affect the temperature calibration when temperatures were measured by the change of melting point of a conductor with pressure. It seemed to be the general opinion of those involved in this and a subsequent discussion that temperatures in synthesis experiments are not known to better than about $\pm 50°$. The scientific importance of these calibrations is that they indicate at which points on the P/T diagram successful syntheses have been achieved in relation to the extrapolated equilibrium line of Berman and Simon.

A successful synthesis depends on having a sufficiently high reaction rate, that is, a low enough free energy barrier, for the transition. Some information about the rate of the direct transition is obtainable from studies of the graphitization of diamond at high pressures and temperatures. Insufficient work has been done to establish these rates accurately and no work has been reported since Bridgman (1947). Professor Neuhaus (Bonn) said that he thought that the equilibrium line at the higher temperatures fell away to pressures much lower than those given by the Berman-Simon line. This was based on a reported zero rate of graphitization (viz. 0·1–0·2%, Bridgman 1947) at an estimated

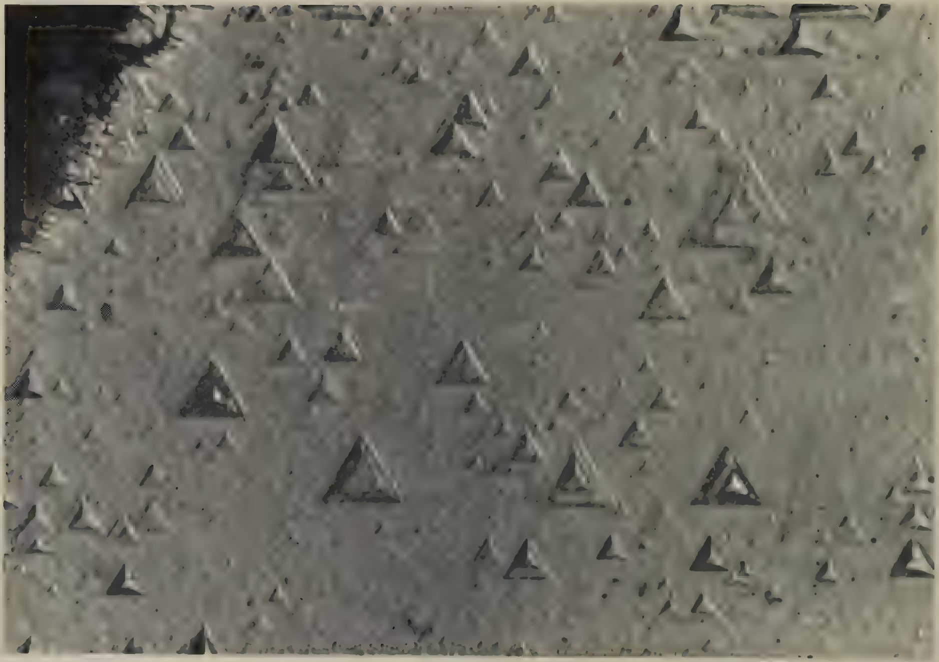


Fig. 1. Triangular etch pits in positive orientation on a (111) diamond surface. The stone had been heated in air for 1/4 hour at 800°c. (×540).
(*from the paper by Evans and Sauter*)

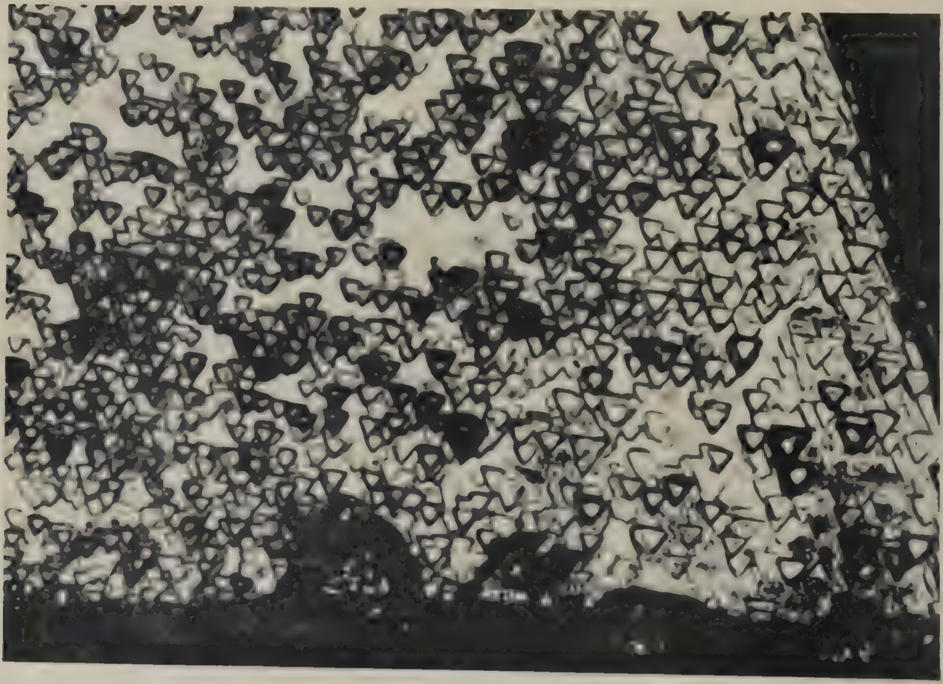


Fig. 2. Triangular etch pits in negative orientation on (111) diamond surface. The stone had been heated in air for 3 minutes at 1350°c. (×192).
(*from the paper by Evans and Sauter*)

3300°c and 30 000 atm. Professor Frank pointed out that this did not necessarily mean that the point was in a diamond-stable region. Professor Neuhaus also showed pictures of natural diamond inclusions which he cliamed were of a non-cubic phase.

The subject of surface graphitization was discussed in papers by Dr. Evans and Mr. Phaal (Reading), Dr. Evans and Mr. Sauter (Reading) and Dr. Seal (U.S.A.). In the paper by Evans and Sauter the topographical effects of surface graphitization and etching was studied. Crystallographically defined pits were produced when diamonds were heated in moist oxygen. Pits were not found when crystals were heated in dry oxygen, H_2O vapour, or mixtures of dry nitrogen and oxygen. The orientation of the pits was a function of temperature, there being a marked change at about 1000°c. Below this temperature, the trigonal pits on (111) were found to be of positive orientation (i.e. the same orientation as the triangular natural octahedral faces), while above 1000°c they were of negative orientation. In the region of 1000°c the pits were hexagonal tending to circular. On the cube faces the square pits were bounded by $\langle 100 \rangle$ when heated at low temperatures, but by $\langle 110 \rangle$ when the crystals were heated above 1000°c. They suggested that the fast moving steps were stabilized at the low temperatures by (C-O-C-O) complexes which were broken down at temperatures above 1000°c.

On the contrary, in his paper Professor Farnsworth (with Dr. Marsh, U.S.A.) suggested that a clear low energy electron diffraction peaks from a (100) surface of a semi-conducting diamond heated for a long period at 500°c in a vacuum (ca. 10^{-9} mm of Hg) indicated a pure carbon surface. The surface spacings which he found were those corresponding with the dimension of the normal unit cube. In the discussion Professor Pryce and Professor Frank wondered whether the clear pattern might not be derived from a diamond surface covered by a monolayer of oxygen. It was considered that this could also give the observed surface spacing.

Experiments on rates of surface oxidation in various ambients were described by Evans and Phaal, while the paper by Seal discussed internal graphitization above 1700°c as well as lower temperature surface effects. Some 3-fold propeller patterns of graphite nuclei were observed in these experiments and similar black patterns were observed by Dr. Howes (Reading) in his high temperature graphitization experiments.

In the Thursday afternoon session some electronic effects were discussed. The first two papers—Professor Ewald (U.S.A.), and Dr. Bolton and Mr. Clark (Newcastle)—described attempts to explain the recent measurements of the intensities of x-ray reflections from diamond by Gottlicher and Wölfel (1959). Dr. Bolton compared these results with those obtained from a calculation based on the Thomas–Fermi method into which corrections were introduced to allow for a tetrahedral angular dependence of the charge density of the carbon atom. Some improvement was obtained. It was also pointed out in discussion by Professor Ewald that in this work it would be valuable to have accurate intensity measurements on many diamonds. Dr. Lomer suggested that a neutron diffraction study would be valuable in assessing the contribution which—in the case of x-rays—the valence electrons make to the (222) reflection.

Also in this session the anisotropy of magneto-resistance in p-type semi-conducting diamond was discussed in a paper by Dr. Mitchell and Mr. Kemmey

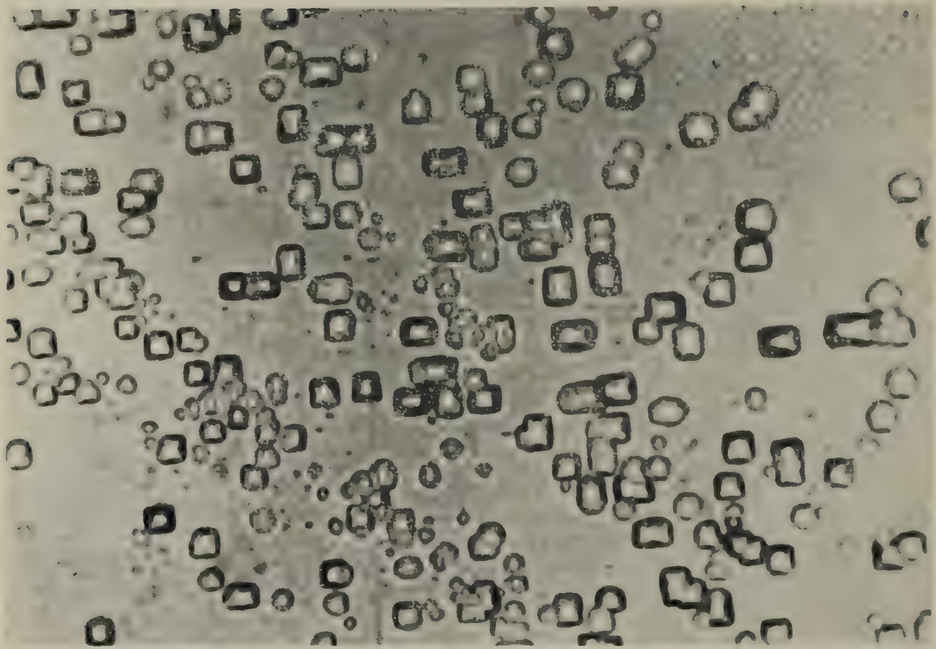


Fig. 3. Square etch pits with sides parallel to ⟨100⟩ directions on a (100) diamond surface. The stone had been heated for 2 hours at 775°c in air at a pressure of 10^{-1} mms. Hg. (× 1680.)

(*from the paper by Evans and Sauter*)

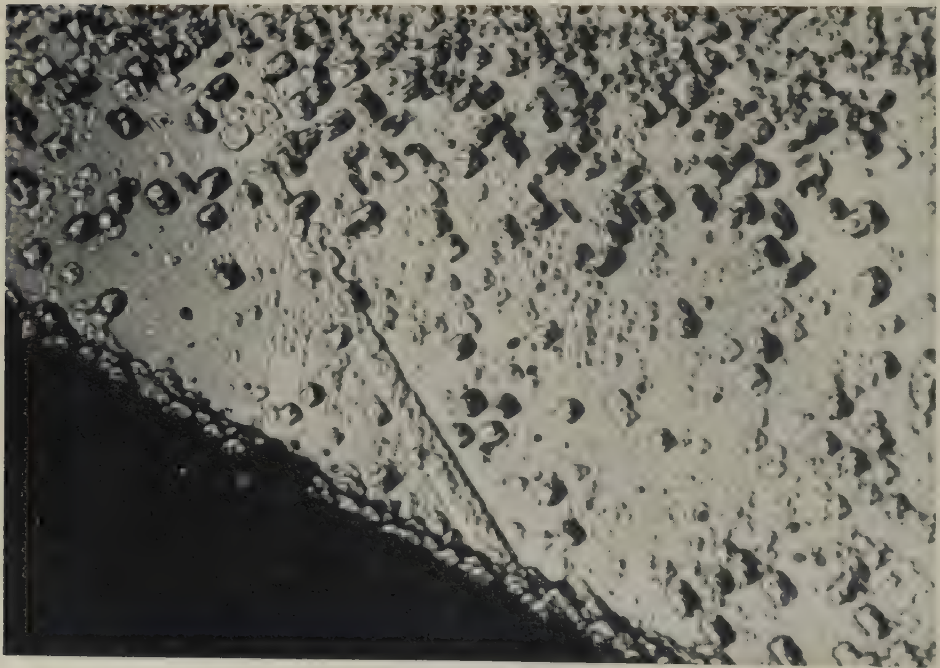


Fig. 4. Square etch pits with sides parallel to ⟨110⟩ directions on a (100) diamond surface. The stone had been heated in air for 6 minutes at 1400°c. (× 540).

(*from the paper by Evans and Sauter*)

(Reading). The existence of the longitudinal effect and the small crystalline anisotropies of all effects indicated that the valence band energy surfaces were not spherical. The three independent components of the magneto-conductivity tensor were calculated from the results and compared with computations based on a model of two warped spheres and one sphere. Using Hall and D.C. conductivity data, the numerical agreement was within a factor of 3, which was considered to be encouraging in view of the various assumptions which had to be made. Dr. R. A. Smith (R.R.E.) drew attention to the importance of uniform impurity distributions in assessing the results of magneto-resistance experiments.

Dr. Clark, Mr. Kemmey and Dr. Mitchell (Reading) described experiments studying the threshold of radiation damage effects in diamond. The rates of introduction of optical absorption in the 1·67 ev band—which gives the bluish green colour— and the rates of increase of resistivity of semiconducting diamond were measured for bombarding electron energies between 0·3 and 2·0 MeV. Below 1 MeV a threshold displacement energy of about 80 ev was indicated, but above 1 MeV the two sets of measurements gave diverging results. The rates determined from the electrical measurements increased rapidly between 1·0 and 2·0 MeV, whereas the rates from the optical measurements were approximately constant between 1·25 and 2·0 MeV.

The effects of heat treatment were described by Dr. Clark and Mr. Palmer (Reading). Damage was annealed in four main stages: 300°c, 500°c, 600–700°c and >700°c. The amount of recovery of optical absorption (1·67 ev) found in the first stage was a function of the irradiation temperature. After this heat treatment the absorption could not be regenerated by x-irradiation and it was suggested that the first stage was an atomic recovery probably involving the recombination of close interstitial-vacancy pairs.

Professor Champion (King's College, London) described measurements of the luminescence exitation spectra near the band-to-band absorption edge. Some of the well known absorption features were resolved, and owing to the greater sensitivity of exitation spectra many features were found which are not detected in absorption. Bands were also observed in the strongly absorbing region. The excitation spectrum from semi-conducting diamond showed considerably less structure than that found for Type I on IIa crystals.

Dr. Smith and Dr. Hardy (Reading) reported that they had produced infrared absorption in a Type IIa diamond irradiated with fast neutrons. This absorption was in the normally forbidden vibrational region and showed a sharp cut-off at the Raman frequency. They also discussed in detail the phonon energy assignments in the 2–phonon absorption region in IIa diamonds (3·75–6μ). As a result of these, Hardy concluded that the dispersion curve of the atomic vibrations was considerably different from that of germanium, the flat portion extending for a smaller distance into the zone from the zone boundary than in the case of germanium. This meant that the long range interatomic forces which have to be considered in germanium were much less important in diamond.

The surface topography of synthetic diamonds from various sources was illustrated by Professor Tolansky (Holloway College, London). Although the crystals were small, Professor Tolansky presented a series of slides of microscopic and interferometric photographs of the surfaces. The crystals from the

G.E. of the U.S.A. frequently contained holes (‘ hopper type ’), the others did not.

In the Friday afternoon session there were three papers devoted to the electronic structure of silicon carbide, SiC, and four papers dealing with friction and wear of silicon carbide and diamond.

The preparation of pure single crystals of SiC was described by Dr. Knippenberg (Holland) and the measurement of their Hall constant and resistivity between 80 and 1400°K discussed by Drs. Knippenberg, van Daal and Wassher. This work illustrated very clearly what serious scientific effort can achieve in a material previously thought to be too difficult. The point was also illustrated by the absorption and luminescence results reports by Drs. Hamilton, Choyke and Patrick (U.S.A.), their paper included an account of the absorption edge measurements in the 4H modification (△△▽▽). The edge was found to exceed that in 6H (△△△▽▽▽) by 0·25 ev. Indirect band to band transitions have been observed in both types and phonon energies of 0·043, 0·069 and 0·094 ev. deduced from the 6H results. They pointed out that it should be possible to provide experimental data on the various types, and that this will provide a very detailed test of the accuracy of band theory calculations.

In the session devoted to Friction and Wear, papers were read by members of the Physics and Chemistry of Solids Group of The Cavendish Laboratory. These dealt with the measurement of hot hardness of diamond and silicon carbide (Dr. Fitzgerald) ; the fracture of diamond and other brittle materials under single and repeated impact (Mr. Field and Mr. Hancox): an apparatus for measuring friction of refractory materials at temperatures up to 2000°C (Mr. C. A. Brookes) ; and the friction and wear of both diamond and silicon carbide when rubbed by steel at high sliding speeds (Dr. Miller).

At a special session on Thursday the Duddell Medal was presented to Professor R. V. Jones (Aberdeen). Professor Jones’ lecture dealt with developments and uses of optical levers, the developments having reached the remarkable stage where the limits are the statistical fluctuations in the light beam.

This was a very successful conference—partly because it was neither too big nor too small. Finally I think that all attending the conference would agree that a considerable contribution to the success of the meeting was made by the local organizing secretary, Dr. C. D. Clark for the thorough but unobtrusive way in which all the arrangements were made.

The Author:

Is Reader in Physics at the University of Reading and directs the research group working on radiation damage effects in solids. The work of the group is primarily concerned with damage produced by fast electrons obtained from the Physics Department’s 2 MeV. Van de Graaff accelerator.

RESEARCH DAY AT THE ROYAL INSTITUTION, NOVEMBER 28TH, 1960

Low Temperature Physics

by D. SHOENBERG
Royal Society Mond Laboratory, Cambridge

The Royal Society Mond Laboratory was built in 1933 to enable Kapitza to expand the work he had pioneered on the properties of matter at very high magnetic fields and to extend it into the realm of very low temperatures. Kapitza's special interests led to an unorthodox but attractive design of building, the central feature of which is a large hall which used to house the high field equipment and around which are grouped the various research rooms which accommodate something like 20 research workers. The high fields were produced by momentarily connecting a large dynamo across a coil, and the length of the hall was such that the mechanical shock at the dynamo (due to its sudden loss of angular momentum) reached the coil and recording equipment only after the recording was completed. The central hall lost its original function when the high field equipment followed Kapitza back to Moscow but its spacious atmosphere is still very pleasant and lessens the isolation of the individual research rooms.

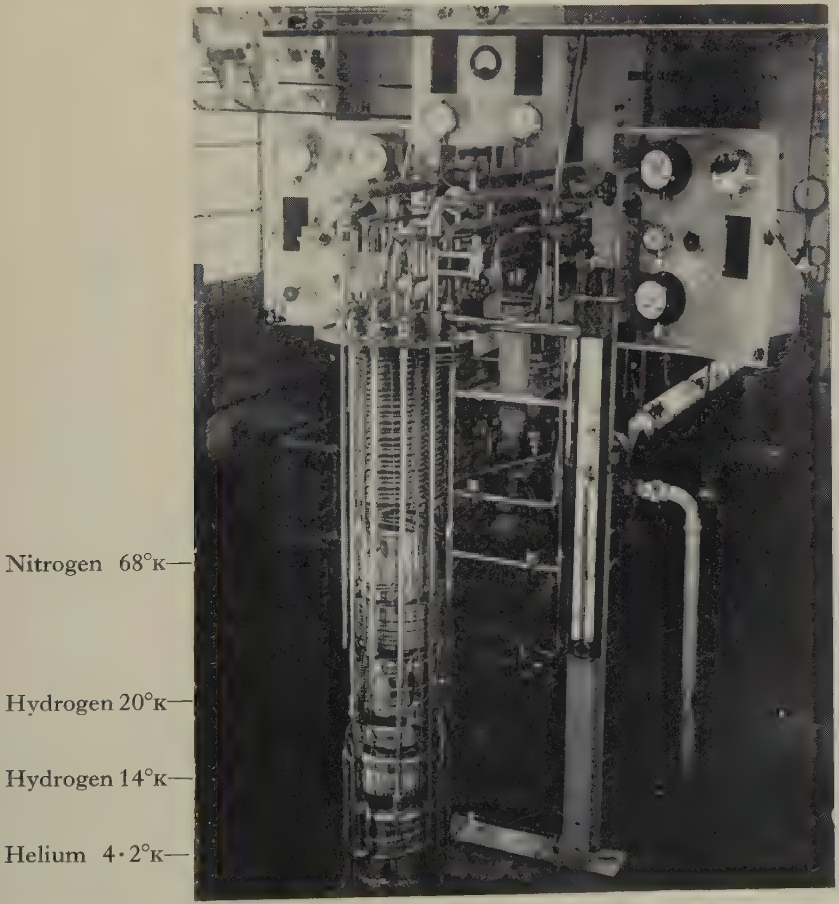


Fig. 1. Ashmead helium liquefier.

1. The Liquid Helium supply

The most important technical facility of the laboratory as it is now is the helium liquefier. This is no longer the adiabatic expansion machine which Kapitza built just before he left in 1934, for although this gave excellent service till after the war it began to be inadequate for the expanding needs of the laboratory and was replaced in 1949 by a cascade type of liquefier (fig. 1) built by Ashmead. This is operated for a few hours on each of two days a week and the normal routine is for the Dewar flask which contains the equipment which needs cooling to low temperatures, to be filled with liquid helium from the liquefier by a siphon tube. The Dewar is then taken back to the research room from where the evaporating helium gas can be collected in a central gas-holder system. Evaporation is greatly reduced by surrounding the helium Dewar with another Dewar containing liquid nitrogen. The temperature of the liquid helium can be reduced from its normal boiling point (4·2°K) to about 1°K, by connecting the Dewar flask to a powerful pump; the temperature can be held at any intermediate level by controlling the rate of pumping.

2. Electrons in metals

One of our main activities in recent years has been a study of the electronic structure of metals, in which certain effects which occur only at very low temperatures are used to give basic information about the individual metals. If the electrons in a metal were completely free, the energy E of an electron would be related to the wave number k of its de Broglie wave by $E = \boldsymbol{h}^2k^2/2m$ (since $k = 1/\lambda = mv/\boldsymbol{h}$). If we imagine a '$k$-space' in which the state of an electron is represented by a point distant k from the origin in the direction of movement of the electron, then a surface of constant energy would be a sphere. According to the Fermi–Dirac statistics of quantum theory the electrons can have energies only up to a certain maximum energy, the Fermi Energy, which depends on the number of electrons per unit volume, and the corresponding constant energy surface in k-space is called the Fermi Surface. In reality this surface would not in general be spherical, because the electrons are not really free but moving in the periodic field of the crystal lattice. The relation between E and k is no longer the simple one for free electrons, because the electron waves are modulated by the periodic field. In theoretical calculations of the properties of metals, it turns out that it is rather vital to know the detailed shape and size of the Fermi Surface and neighbouring constant energy surfaces. Although in principle this information, which we call a little loosely the 'electronic structure', could be calculated theoretically from a knowledge of the structure of the atoms and their arrangement in the crystal lattice, in practice the calculation cannot be carried through reliably enough and we must fall back on experiments.

There are a number of experimental methods but I shall today speak only of two: the de Haas–van Alphen effect and cyclotron resonance. The de Haas–van Alphen effect is an oscillatory variation of the feeble diamagnetism of a metal single crystal which occurs at low temperatures when the magnetic field is varied. This effect comes about essentially because of the quantization of the orbits into which the free electrons are bent by the magnetic field and the theory shows that the 'interval in reciprocal field' between one oscillation and the next, i.e. the period P of the de Haas–van Alphen oscillations (fig. 2), is inversely proportional to the extreme area of cross section of the Fermi Surface in a plane normal to the field direction. Thus by varying the field direction relative to the crystal axes we can get a lot of information about the size and shape of the Fermi surface. The reason that this effect is seen only at very low temperatures is that kT (k is here Boltzmann's constant) measures the ' smearing ' of the boundary between occupied and unoccupied states and this must not be too large compared with the energy separation of successive quantized orbits, if the amplitude of the oscillations is not to become unobservably small. This separation for a field H is $hHe/2\pi m$ (e is the electronic charge, h is Planck's constant and m is an effective electronic mass) and when numbers are put in we find that to get a reasonable amplitude we not only have to have T as low as possible, say 1°K ,but H as large as possible. The field of an ordinary electromagnet say 20 000 gauss. is not large enough except for rather anomalous electronic structures where m may be appreciably smaller than the true electron mass. For a metal like copper m is fairly normal and fields of 100 000 gauss or more are needed.

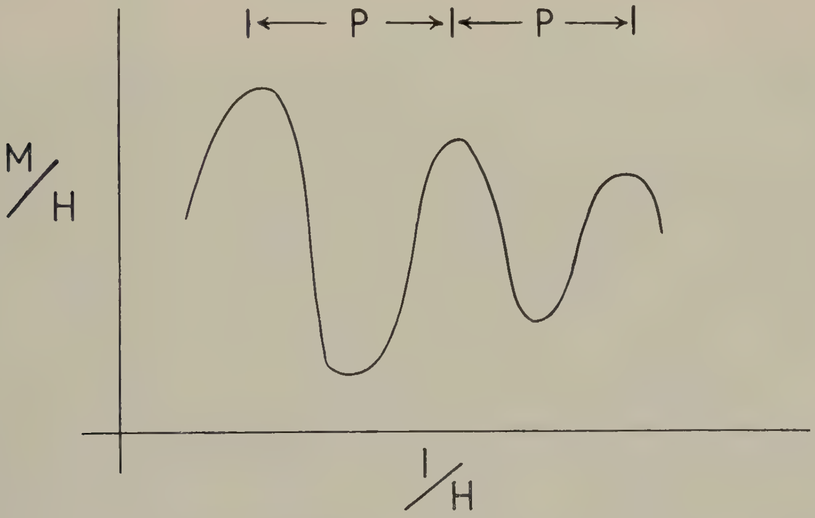


Fig. 2. Plot of M/H, representing the susceptibility, against 1/H for de Haas—van Alphen oscillations. This shows schematically the information *derived* from the time-oscillogram of Fig. 3. The "period" P in intervals of 1/H is the information needed to explore the Fermi Surface.

It is because of this that we have revived the Kapitza tradition of high magnetic fields. The difficulty about generating high magnetic fields is to get rid of the heat produced by the necessary high currents. Kapitza overcame the difficulty by producing the high fields for only a short time, and as I have already said, this he did by momentarily connecting a large dynamo across a coil. Nowadays it is easier to store the necessary energy in a bank of condensers rather than in the rotation of the dynamo, and in fact we get 200 000 gauss by discharging 4000 μF charged to 2500 v through a suitable coil cooled in liquid air. The time scale is such that the peak field is reached 7 ms after the start of the discharge. To detect the de Haas–van Alphen oscillations the metal single crystal is surrounded by a pick-up coil and the e.m.f. produced by the varying magnetic moment of the specimen as the field varies is amplified and displayed on a C.R.O. (fig. 3).

Another approach to the electron structure of metals is by way of cyclotron resonance. I have already mentioned that in a magnetic field H the quantum levels are separated by energy steps of $hHe/2\pi m$. If now a high frequency radio wave impinges on the metal of frequency ν such that $h\nu$ is just equal to this separation or to a multiple of it, i.e. if $\nu = nhHe/2\pi m$ (n is an integer) then this wave will suffer absorption, and we speak of cyclotron resonance, because the basic frequency involved is just that of a cyclotron. This frequency comes out of order 30 000 mc/sec if H is 10 000 gauss, i.e. radio waves of 1 cm wavelength, and the experiment is actually done by observing how the Q of a resonator of given frequency varies when the field changes. You will notice that this experiment does not really tell us about the Fermi Surface itself but gives only a value of the effective mass. It can be shown that this effective mass is determined by the rate at which the area of cross section of a constant energy surface changes with energy just at the Fermi Surface. The low temperature is necessary here simply to ensure that the mean free path of the electrons is long compared with the circumference of the orbit into which it is bent by the field, for otherwise the energy levels would become blurred and the resonance would not be sharp.

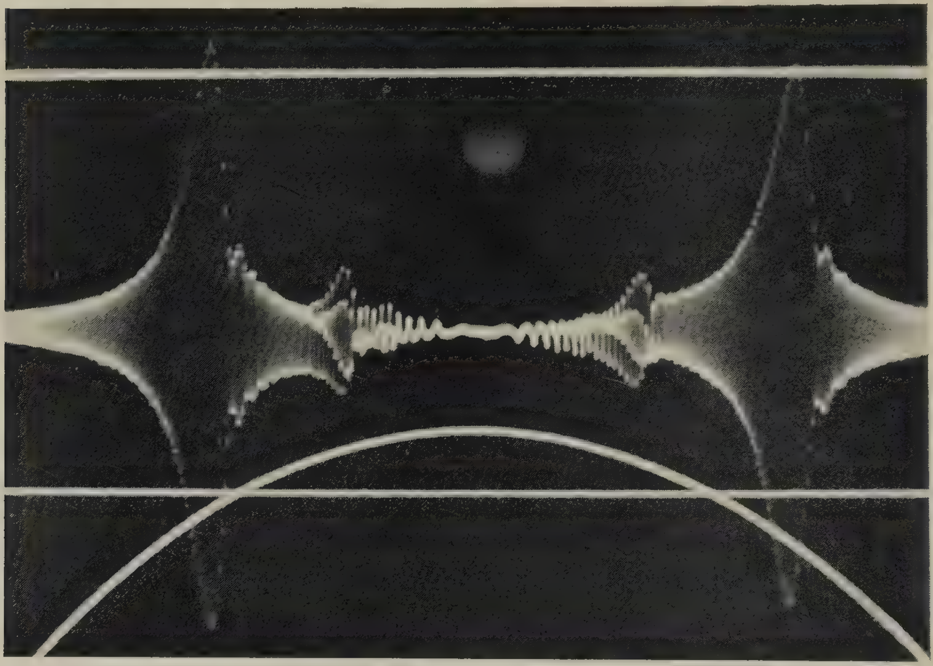


Fig. 3. Oscillogram of de Haas-van Alphen oscillations in a copper single crystal. This shows the variation *in time* of M while H, which is indicated by the parabolic trace, is varying *in time*. The time-period of these oscillations increases as dH/dt decreases.

3. LIQUID HELIUM

Let me turn now to quite another low temperature problem, that of liquid helium itself. The isotopes of mass 4 (He^4, the common one) and of mass 3 (He^3, the rare one, produced mostly as a by-product of the American defence programme) behave quite differently. Both need low temperatures for liquefaction because of the feebleness of the forces between helium atoms, but the liquefaction is also considerably influenced by a quantum effect known as zero-point energy. There is in fact considerable energy of agitation of the atoms even at absolute zero, and since this energy is larger for the lighter atom, it is harder to liquefy helium 3 than helium 4 (the boiling points are 3·2°K and 4·2°K respectively). It is because of zero-point energy too that neither species of helium solidifies except under pressure.

Until recently most research was concentrated on liquid helium 4. This shows a remarkable change at 2·2°K (the λ-point) ; below this temperature it becomes such an excellent conductor of heat that no bubbles can form and the λ-point transition is shown by the apparent sudden cessation of boiling (evaporation continues, of course, but only from the top surface). It also becomes superfluid, so that no pressure is needed to drive the fluid through very narrow channels, though experiments show that the liquid exerts an appreciable damping on moving objects in it as if it still had viscosity. Another remarkable phenomenon is the 'fountain effect' discovered by Allen in our lab. just before the war. When heat is applied to closely packed emery powder at the bottom

Fig. 4. The "fountain effect" in liquid helium.

of an open tube in the liquid, the liquid seems to be attracted to the warmth so strongly that a fountain emerges from the open end. These and many other curious effects have been explained in terms of a ‘ two-fluid ’ theory according to which, as the temperature is lowered below 2·2°K, an increasing proportion of the fluid goes into a quantum state of superfluidity, while the remaining part in which the thermal motion is concentrated remains ‘ normal ’ (and is responsible for the friction). One interesting prediction of this theory is that if liquid helium is rotated the rotational angular momentum of the superfluid is concentrated in quantized vortex lines. Recently Dr. Vinen has been able to demonstrate that this is really so by an ingenious experiment and he will tell you about this later.

Because it is felt that liquid helium 4 is now more or less understood in principle, we are turning our attention to liquid helium 3. The important difference between the two isotopes is not just that the masses are different but that He^3 atoms have a nuclear spin and so (like electrons) obey Fermi–Dirac statistics, while He^4 atoms have no nuclear spin and so obey Bose–Einstein statistics. The theoretical problem is however greatly complicated by the close proximity of the atoms to each other in the liquid state and it is by no means certain how liquid helium 3 should behave at very low temperatures. So far it has been studied experimentally only down to 0·1°K and it shows none of the remarkable features of liquid helium 4. It is however possible that something peculiar might happen at still lower temperatures. For instance, it is possible that a transition to superfluidity might show up as in helium 4, while according to another theory there should be a very strong absorption of sound at sufficiently low temperatures. Dr. Abraham, Mr. Gough and Dr. Vinen have set up an elaborate equipment to study the propagation of sound down to 0·01°K. This project combines nearly all the ingredients of the low temperature physicist’s bag of tricks—liquid helium 4, liquid helium 3, adiabatic demagnetization and of course lots of ‘ plumbing ’ and electronics.

4. Superconductivity

Perhaps the best known characteristically low temperature phenomenon is superconductivity, and a good deal of our effort has been devoted to its study. Certain metals such as tin, lead, etc., when cooled below a ‘ transition ’ temperature characteristic of the particular metal abruptly lose all electrical resistance and are then said to be superconducting. I will now demonstrate that a superconductor has remarkable magnetic properties. Sitting at the bottom of the liquid helium in the inner Dewar flask is a tin rod surrounded by a coil; the liquid helium is at 4·2°K and the tin is not yet superconducting. I now switch on a small magnetic field and the kick of the galvanometer measures the magnetic flux which has entered the tin rod (the rest of the flux is compensated by an extra coil). I now cool the tin rapidly by pumping on the liquid helium and you will see that as the tin passes through its transition temperature at about 3·7°K the galvanometer kicks the opposite way indicating that the flux is ejected when the tin becomes superconducting. This illustrates the famous Meissner effect, i.e. that a superconductor cannot tolerate any magnetic flux inside itself and so ejects the flux when it becomes superconducting. If I now increase the magnetic field (notice that there is no galvanometer kick at first because no flux enters) a point is reached when the superconductivity is killed, i.e. the 'critical' field has been reached. The flux now enters again as evidenced by a new kick, and when I reduce the field again we again see an opposite kick as soon as superconductivity is re-established. The Meissner effect means that a superconductor is *very strongly* diamagnetic and I have another demonstration, which shows a small magnet floating above a superconducting lead dish; the diamagnetism of the lead repels the magnet strongly enough to overcome gravity.

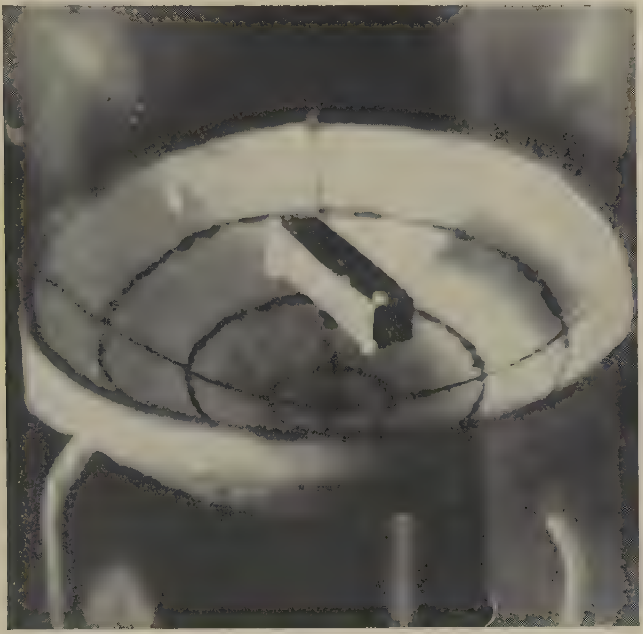


Fig. 5. Illustration of the Meissner effect; a small magnet poised over a superconducting lead dish. [*Courtesy of Cambridge University Press*].

The circulating currents which cause this diamagnetism are rather like the quantum orbits in an atom and what you have seen is essentially a manifestation of a quantum effect on a macroscopic scale. As for the case of liquid helium we have to think of only some of the electrons as being in the superconducting state while others (a fraction which diminishes as the temperature is lowered) are in excited states and behave as if they were still 'normal' and would be resistive if they could be stirred up. The 'normal' and 'superconducting' electrons behave very much like a pure resistance and a pure inductance in parallel. For D.C. the inductance carries all the current and there is no resistance, but for high frequency A.C. the impedance of the inductance becomes comparable to that of the resistance and a good deal of the current is carried by the resistance (i.e. by the 'normal' electrons) (fig. 6). Much of our research on superconductivity has been concerned with the behaviour of superconductors at very high frequencies and Dr. Adkins is doing experiments at 150 000 mc/sec (2 mm waves). He finds not only the resistive effect of the normal electrons but evidence for transitions of the superconducting electrons over the energy gap which separates them from the excited states. Such transitions happen when the *hν* of his waves is greater than the energy gap and his experiment is designed to check some of the finer details of the theory.

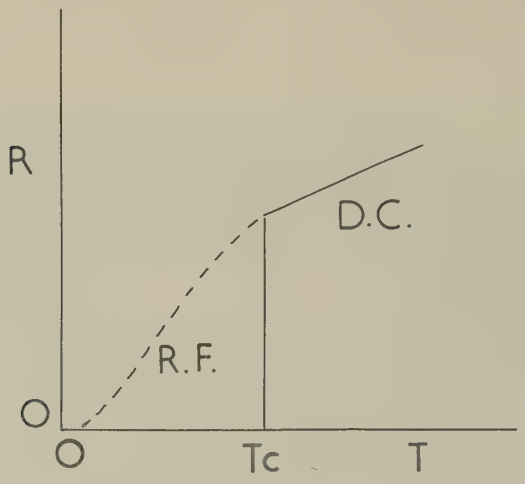


Fig. 6. Variation of resistance with temperature in neighbourhood of absolute zero (schematic). The full line is for direct current, the broken line for radio-frequency alternating current.

Suggestions for further reading:

Atkins, K. R., *Liquid Helium* (Cambridge University Press), 1959.
Mendelssohn, K. A. G., 1960, *Cryophysics* (Interscience Publishers).
Pippard, A. B., 1960, Experimental Analysis of the Electronic Structure of Metals, *Rep. Prog. Phys.*, **23**, 176, 1960.
Shoenberg, D., *Superconductivity* (Cambridge University Press), 1960.

The Author:

Dr. David Schoenberg was educated at Latymer Upper School, W.6. and Trinity College, Cambridge. He started work at the then newly built Royal Society Mond Laboratory under Kapitza in 1932 and has been in charge of the Laboratory since 1947. His own researches have been mainly concerned with superconductivity and with the de Haas-van Alphen effect. He is a Fellow of Gonville and Caius College and was elected F.R.S. in 1953.

Contemporary Reading

This is not a full survey or review of the two official Reports that it deals with, but a discussion of some of the important points that they have in common. One Report, on sixth form examinations in general, applies to the teaching of physics in that a high proportion of the candidates are taking physics as one subject; the other, concerning itself with the national scientific policy, spreads itself on the educational side of this.

My only difficulty with both of them is to interpret two terms that repeatedly occur. These have acquired such strong emotional content that whenever I see them I wonder if I am being got at, and if the user has been nobbled in his turn. The first is ' factual knowledge ', or ' factual content ', the only definition of which seems to be that it is there to be cut down. The second is ' specialization ', which is excessive, overdone, harmful, and (*vide* L. Putt's opuscule *The Specialist*, published by Blackwell's) of ill-odour. They are hardly objective terms any longer. It could be argued that pupils do not nowadays acquire *enough* factual knowledge in relation to the time spent on scientific studies, and that the habit of hard work has steadily declined during the past twenty years. And that genuine specialization, in the large number of schools where it is successfully practised, seems to produce lively people whose interests overflow in all directions. Sixth-form work sorts out the ability-spectrum very effectively; and what is basically wrong with things here is that with large sixth forms it is impossible to let pupils specialize as individuals.

Now, education in science is not a particularly prosperous concern at the moment. It is like a business that is floundering because it hasn't the staff or the capital to handle a vast new demand for an up-to-date model of its product, and has to take steps within the framework of what is politically possible. It might

(*a*) merely call in the auditors to reorganize its book-keeping;

(*b*) produce a slightly cheaper popular model of the same old job, or

(*c*) re-tool the whole production line, re-train the staff in its use, and look ahead to the time when it can attract fresh staff and capital to an enterprising firm.

The General Certificate of Education and Sixth Form Studies, the Third Report of the Secondary Schools Examinations Council (H.M.S.O., 1960, 2*s.* 6*d.*) discusses all three measures. For (*a*) it suggests a modification of the examinations and the method of presenting the results, and also the use to be made of them by universities. Under (*b*), a reduction in the factual content of the syllabuses, which is admirable in intent until one thinks about what it means for Advanced-level physics. For the basic factual content is taught several years earlier; the advanced work done at present consists largely of analytical development and applications. And under (*c*), it proposes that a concerted attempt should be made to revise syllabuses. This is the line suggested by the document from the Office for Scientific and Technical Personnel of the Organization for European Economic Cooperation, entitled *The Teaching of Physics in Schools*, which outlines the principles on which advanced courses both for the non-specialist and the future specialist (note, *future*, not *harmed*) might be based. It also indicates lines along which the technique (as distinct from the book-keeping) of examinations might be improved, some of which agree closely with those of the SSEC Report—particularly in the matter of testing knowledge and understanding rather than memorized information.

The Annual Report of the Advisory Council on Scientific Policy 1959–60 (Cmnd 1167, H.M.S.O., 1960, 1*s*. 9*d*.) will be quoted verbatim. This is what it says:

> *School Science Curricula*
>
> 46. We have no doubt that school science curricula are in need of a thorough re-examination. They tend at present to be unimaginative and to be overloaded with factual material (in part as a result of the tendency to keep adding new material without removing the old). It has been suggested to us that up to 20 or 25 per cent of the content of the curricula in physics, chemistry and biology could be removed without any harm—and indeed with benefit. Mathematics curricula are equally in need of revision.
>
> 47. In particular there is a need for a new approach to the teaching of biology in schools. Biology is particularly well suited as an introduction to science, but is often taught as an incompletely fused amalgam of classical botany and classical zoology. It should instead be treated as a comprehensive discipline in its own right, with a bearing on the teaching of geology and geography. It should involve much more field work than is customary at present. In this revised form biology should be an integral part of the junior and middle school curriculum for all children.
>
> 48. Extensive studies on the content of school science courses are now in progress in the United States under the auspices of the National Science Foundation, and in physics much material is now available, including new text books and films. The Council understand that the Ministry of Education has established contact with the National Science Foundation on this aspect, and hope that the closest possible contact will be maintained.
>
> *Specialisation*
>
> 49. We consider that specialisation in schools has gone too far, and that in it originates the much discussed ‘ gulf ’ between scientists and non-scientists. We deplore the segregation of children in the middle school into groups according to the subjects they are to study in the sixth form, and believe that, with the revision of science and mathematics curricula mentioned above, it should be possible for all middle school pupils to study science without making undue inroads into other subjects. In our opinion all pupils in the middle school should have the same general education, in which science should play its full part.
>
> 50. While we appreciate the desire of teachers and pupils to pursue some subject in depth, we believe that the degree of specialisation usual today is excessive, and that the pressure on both sides—arts and science—can be relieved by modifications of curricula so that those who are primarily interested in science should continue a serious study of some other subjects and vice versa. We do not regard a few periods of general studies sufficient as a leavening for the Science Sixth, nor a few periods of general science for the others.
>
> 51. Even within science there are signs of over-specialisation in the sixth form. This is especially true of biologists; it would be greatly to their advantage to learn mathematics, physics, and chemistry throughout their school period, even if it meant some reduction in the amount of biology studies.
>
> 52. As things stand at present, we are bound to agree with the Crowther Committee’s view that two science subjects may be all that a student should take at the G.C.E. Advanced Level. But this is true only if the extent of factual knowledge in individual subjects demanded at A-level remains unchanged. We believe that this should be reduced so that, with curricula revised and reduced in the manner indicated, the breadth of sixth form education can be correspondingly increased.

53. An important reason for over-specialisation in sixth forms is the nature of College Scholarship requirements at Oxford and Cambridge (as distinct from entry requirements). Although only a comparatively small number of pupils take these Scholarship examinations, the standard set has a quite disproportionate effect on school curricula. There is an urgent need for schools and universities to get together and see what can be done to devise some method of awarding scholarships which will avoid this.

54. Attention was drawn in our discussions to the generally recognised fact that there is considerable overlap between what a student is taught in science in his first year at university and what he has already learned in his last year at school. Many professors would welcome a lower level of factual knowledge of their own subject in pupils entering the university, if it were combined with a broader general education. It should be remembered that such a combination of less specialised factual knowledge with a greater range of subjects has long been the rule, not only in Germany and the United States, but also in Scotland.

Many of these points were made also in the SSEC Report, and the recommendations on the whole seem generally desirable. In particular, paragraph 48 is welcome; we have tended to become complacent about our own ideas and standards. It might be mentioned here that the physics material referred to is that designed for a *one year's course* in physics for pupils between 15 and 17 who have had no previous background, and that a good deal of it (though not all) would be quite suitable for use in sixth-form work.

Specialization here means ' over-consumption of time ', both in the middle school and in the sixth form. There is a law (I've forgotten its name) that says ' work expands to fill the time available '. We certainly spend too much time at present on stereotyped practical work (about which the OEEC paper has plenty to say) preparing for a stereotyped practical examination (which the OEEC paper would see abolished in all its forms). And it must not be forgotten that much physics teaching is geared to a limited stock of aged equipment, which there is no technician to service or extend, and which wastes time and effort. Let us suppose that a new and enlightened syllabus is agreed on, and that this is allowed only the time required to cover it efficiently. What does this really entail ? Better equipment, competent assistance, better libraries, better books, the full resources of all modern teaching aids, more opportunities for teachers to study the presentation of their subject, and more time for them to reflect and prepare their work. It is not just a matter of saying ' this is bad and that is good '; active planning to provide the right conditions in which the good can flourish, and action with man-power and money behind it, are needed.

On paragraph 53, it is true that the present system is swamped by numbers, and blessed chiefly by tradition. At the College end, there is confusion (in variety) between competition for a scholarship or for a place; the chief complaint is this particular untidiness. The rational answer would seem to be to select people for places *earlier* (a schoolmaster can certainly sort them out after a term in the sixth form, and spot potential scholars much sooner), and to confine competition for scholarships to those who have already been accepted for a place, or are actually in residence, as Sir John Cockcroft has recently suggested. (The Group scheme in which Churchill College is a leading participant, is a move towards tidying things up at Cambridge.) Now, if the system has defects, it does not follow that it is an evil influence on a national scale. This criticism amounts to saying that the Derby and the St. Leger are the ruination of racing because they attract non-starters, runners doped with stimulants, and horses that can't stay the distance. If they were, I can't help feeling that the Jockey Club would start on the trainers rather than on the courses; but important matters are handled differently from education.

But many of the opponents of the scholarship examination system are not concerned so much with any possible effect on curricula as with a process that appears to give Oxford and Cambridge the pick of the most promising. Of this, it can only be said that, if open scholarships were abolished tomorrow, any thwarted scholar would presumably be able to qualify for entrance. And national enterprises and many industrial concerns (I counted 34 on a recent list) are in the scholarship-awarding business; though it must be admitted that they expect a candidate to have secured a place at a College first in many cases. I do not see what can ever be done by any curriculum or examination reform as long as huge numbers of people put Oxford or Cambridge as their first choice, and these institutions take first pick. If this is not in the best interests of the country's educational system, what can be done about it anyhow?

Paragraph 54, with the implication that a student is ' taught ' at university, but ' learns ' at school, has inadvertently diagnosed something. But schoolmasters cannot be complacent, for the usual complaint from the seats of teaching is that students do not ' know ' what they have ' done ' at school; *weakness in elementary facts* is the standard criticism. A quick look at the *Commonwealth Universities Yearbook* shows that, as a proportion of the total professoriate, there cannot be ' many ' professors who could count on any acquaintance at all with their subject in people coming straight from school. From the other side, not more than 5 per cent of those specializing in physics in the sixth form go on to read for a physics degree at the university. It looks as if physics is, in fact, part of the general education of the future scientist already; and this is one of the strongest arguments for enlightened syllabuses which emphasize its educational importance. And another definition of specialization in the bad sense emerges—the anticipation of university work; yes, this is wasteful, inefficient, and *unnecessary*.

I have fumbled through this with the guilty conscience of a man who has spent thirty years instilling factual knowledge into specialists. On his own definition of both terms. But there is another term that I should like to see defined. ' Broad general education '.

There is a feeling of uneasiness about tendencies to lower standards, about the underestimate of what school pupils can rise to if they are inspired and extended, and about the doubtful assumption that everyone can be a first-class cricketer if the boundaries are made small enough and the bowling sufficiently wide and general—instead of all that old-fashioned nonsense about sweating away at the nets.

We cannot lower *standards*. What we can do is to choose appropriate work, at a high enough standard, selected from a wider spread of subjects, that will not restrict a pupil's future choice of studies; and this is what general education must mean.

I recently attended the second of Mr. A. D. C. Peterson's Conferences on Arts and Science in the Sixth Form at Oxford. The Peterson proposals of four Advanced Level subjects as outlined in the Gulbenkian Foundation Report were set beside those of the Headmaster of Winchester, Mr. H. D. P. Lee, for a grouping of ' coherent ' studies to the extent of 2½ subjects, with 1½ non-specialist subjects. The two main lines of agreement were that general education should continue throughout schooldays, and that steps to ensure this should be taken quickly.

The proposals of the Science Masters' Association for new syllabuses at Ordinary-level and Advanced-level have just been published, as these proofs are going to the press, and it is hoped and expected that they will be accepted and acted upon. For they do get down to realities, and speak with a firm decisiveness of aim.

When people are agreed about what is required of a specialist study at school, what is meant by ' factual knowledge ', and what a general education at school implies in these days, they should be well on the way to supporting the necessary quick and energetic steps.

G. R. N.

Optics at Imperial College

The occupation of new optical laboratories and workshops in the new Physics Building marks a fresh stage in the history of the Optics Section under Professor W. D. Wright, and inaugurates the new Optical Design Group directed by Dr. C. G. Wynne.

Specialist instruction in optics has been given at the College since the Department of Optical Engineering was established in 1917, but in 1931 the status of the department was changed to that of a Postgraduate Section of the Physics Department. As a school of optics, the Section occupies a position comparable with the Institut d'Optique in Paris, the Institute of Optics, Rochester, U.S.A., the State Optical Institute, Leningrad, the Institute of Optical Research, Stockholm, the Optisches Institut, Berlin, the Istituto Nazionale di Ottica, Florence, the Instituto de Optica, Madrid, and the optical laboratories of the Institut voor technische Physica, Delft.

Distinguished contributions to optics and optical instrumentation have been made by former professors in the Section—F. J. Cheshire, A. E. Conrady, L. C. Martin and A. F. C. Pollard.

Close relationship with the optical industry is maintained, notably through the Technical Optics Advisory Committee, with its members drawn from industry, the National Physical Laboratory, the British Scientific Instrument Research Association and other bodies.

The 45th Annual Exhibition of The Institute of Physics and The Physical Society

At the Royal Horticultural Society's Halls, 16–21 January 1961

The excellently produced handbook of the Exhibition will serve as a record for those who were fortunate enough to visit it, and can be recommended as a splendid detailed account for those who did not. Even the briefest comprehensive survey would approach its 300 pages in length, so that this report will mention only those exhibits which seemed particularly interesting, as new applications of fundamental physics. This is a rather narrow approach; but space, rather than personal inclination, is the decisive factor.

Whenever I hear the words ‘ two cultures ’ I reach for my double-barrelled gun; and it really is time we disowned the fallacies of yesteryear. The Research Laboratory for Archaeology and the History of Art at Oxford illustrated this point, with its impressive display of x-ray and β-ray back-scatter devices for the non-destructive testing of works of art. The proton gradiometer for measuring small changes in the ambient magnetic field is a new tool that locates largish items of buried pottery by virtue of the thermo-remanent magnetism in baked clay, and can also detect the relics of soil cultivation, which converts the soil's Fe_2O_3 from the weakly magnetic α- to the highly magnetic γ-form. I just do not believe that any scientist could fail to be thrilled by this typical excursion into the realm of humane studies, or that any humanist capable of distinguishing storied urn from animated bust could fail to understand its principles. Culture is one and indivisible; but it appears fragmented nowadays because we are incited to look at it that way.

And what of the other more technical exhibits? First, the Esaki or ‘ tunnel ’ diode, of which several examples were on view. This uses the quantum-mechanical effect of tunnelling through a potential barrier at the p–n junction in a germanium antimonide or indium antimonide junction diode, and operates at very high speed; switching, memory and logic elements, and amplitude–time discrimination are among its uses. Thin ferromagnetic films, in which the domain directions can be diverted (and visibly examined using the magneto-optical Kerr effect) by pulsed signals were another range of new computer element devices. To the physicist, elegant applications of basic principles which work in nanoseconds; to others, a further step towards the revolution of our commerce and organization by fast versatile computers of manageable size.

Thermonuclear research was featured in a large display by Associated Electrical Insutries of the instrumentation for a torus of the Sceptre type, shown operating on a gas discharge tube. Apparatus for the production of 14 MeV neutrons by the (D, T) reaction, in which a tritium-loaded target bombarded by 100 kv deuterons was the source, indicated the interest in studying the behaviour of neutrons of this energy, which are one of the chief ways in which energy is carried away from the plasma in a thermonuclear device, and presumably one of the major problems to be handled. There were three such generators on show, yielding neutron fluxes of the order 10^8 neutrons per second.

Low-noise microwave amplifiers were represented in a variety of types. Solid-state masers; parametric amplifiers, in one case illustrated by a mechanical model in which a driver (signal) pendulum was coupled to an energized (idler) pendulum, the lengths of both being periodically varied—which drew from a companion the remark that this was not a model at all, it *was* a parametric amplifier; travelling-wave tubes, which seem to be the really practicable low-noise amplifiers for satellite-relay developments; backward-wave amplifiers; a cyclotron wave amplifier which operates by increasing the rotational energy of electrons passing down the tube; and an elaborate quadrupole cyclotron wave amplifier, again from AEI, which seemed from the details provided to be a sort of amalgamation of the structure of an atomic-beam apparatus, the electronics of a klystron, and the microwavery of cyclotron resonance. Here again is one side of a partnership; for whom are the efforts at improved communications (which much of the work represents) being made, if not for those who have something to communicate?

From somewhere in Faraday Avenue (the most appropriate of all locations) a brilliant flashing lighthouse beam swung round the New Hall; this was not catalogued, was not really obvious when one was close to the GEC (or AEI) display, and I can only say that I understood it to be a krypton arc. The iodine-containing high-temperature tungsten lamp, and its combination with a mercury arc, were certainly GEC's.

It does not seem so very long ago since electron-spin was, to the ordinary physicist at any rate, an esoteric proposition for experiment; we knew it was there, for theory and the spectroscopists said so; and we should have been hard up without it, because the exclusion principle needed it. But the microwave era has changed all that, since in a magnetic field electrons of opposite spins develop energy differences corresponding to quanta of microwave frequency. This is the basis of the electron-spin resonance spectrometer shown by Hilger and Watts. A more elaborate direct application of electron-spin energy differences was the rubidium-vapour magnetometer shown by the Signals Research and Development Establishment of the Ministry of Aviation. The magnetic field H to be measured causes a microwave energy separation nearly proportional to H between adjacent spin levels; only the lower of these can absorb light circularly polarized in a given sense, so that when the tube of rubidium vapour is irradiated with circularly polarized rubidium D_1 light, this level is depopulated. Selection rules demand that the electron returns to the higher of the two levels, which is thus overpopulated, and absorption stops. To return the electrons from there to the lower level, so that absorption can continue, stimulated emission along maser lines is used—but by an alternating magnetic field normal to H of frequency corresponding to the spin-level energy difference. So that H is determined by finding the frequency of the alternating field at which resonance occurs.

AERE demonstrated the Mossbauer effect to show the Doppler change in the sharp γ-ray resonance of ^{57}Fe when the absorber target of ^{57}Fe moves relatively to the source. This was the resonance used early in 1960 to detect the frequency change of γ-radiation due to the earth's gravitational field, the ' gravitational red-shift ' of general relativity.

The general impression I brought away was that the exhibitors had spread themselves handsomely in order to stage working demonstrations and to make them reasonably intelligible to the ordinary visitor. And it was good to see the contributions from the Science Museum, from some University Departments, and the prominently arranged stand of the Institute and Society which dispensed information about its activities and showed its wide range of publications.

G. R. N.

The Science Masters' Association Annual Meeting

Glasgow, 3rd to 6th January, 1961.

There was an attendance of about four hundred and fifty members at the Annual Meeting of the Science Masters' Association, held in the University of Glasgow from Tuesday, 3rd January to Friday, 6th January, 1961. The Association was very glad to welcome members from as far afield as Rhodesia and Nigeria, and to strengthen the bonds between the Association and similar organisations in the United States through the presence of Dr. Francis L. Friedman, who came over to discuss the work of the American Physical Sciences Study Committee, and five representatives of the U.S. National Science Teachers Association.

The formal proceedings opened on the Tuesday evening with the Science and Citizenship Lecture, on 'The Human Sciences in a Scientific World', given in the Bute Hall by Mr. J. A. Mack, M.A., Stevenson Lecturer in Citizenship. On Wednesday evening, January 4th, the Association Dinner was followed by the Presidential Address by Lord Boyd-Orr, whose subject was 'Education for the Atomic Age'. After Dinner on Thursday, January 5th, the Association was honoured by a University Reception in the Bute Hall, where the guests were received by the Principal of the University, Sir Hector Hetherington. After the Reception, about three hundred and fifty members took the opportunity to visit the Hunterian Museum in its impressive setting. A civic welcome was given to members on Friday, January 6, when they were received by the Lord Provost of Glasgow.

A full programme of lectures had been arranged for the morning sessions. These included

(1) '*Big Rings*' by Professor R. A. Raphael, D.Sc., Regius Professor of Chemistry ; a review of recent work on the novel and somewhat unexpected properties of systems involving rings larger than the normal five- and six-membered rings in organic cyclic structures.

(2) '*The Detection and Measurement of Elementary Particles and Quanta*' by Professor P. I. Dee, C.B.E., F.R.S., Regius Professor of Natural Philosophy ; a development and demonstration of some recent techniques for the detection and quantitative measurement of individual particles and quanta.

(3) '*Better Teaching of Zoology*' by Dr. Hugh Steedman, Senior Lecturer in Zoology; a review of the principles which underlie the rubber latex injection method for demonstrating anatomical detail and of modern methods for the preservation of animal material based on the use of plastics.
We were very grateful to Dr. Steedman for stepping into the breach when Professor C. M. Yonge was unable through indisposition to give his advertised lecture on *Association (Symbiosis) between Invertebrates and Unicellular Algae*'.

(4) '*The Boundaries of Science*' by Dr. Magnus Pyke, B.Sc., Ph.D., F.R.S.E., Glenochil Research Station, The Distillers Company Ltd.; a review of the arbitrary and unreal nature of the boundaries between the sciences, and the limitations of the scientific way of assessing truths.

(5) '*Human Physiology as a Biological Subject in Schools*' by Professor R. C. Garry, M.B., Ch.B., D.Sc., Regius Professor of Physiology ; Human Physiology, an integral part of any course in Biology, lends itself to practical work, and many fundamental facts about the nervous system and the senses can be examined with relatively simple equipment.

(6) '*The Study of Molecules by X-ray Crystal Structure Analysis*' by Dr. J. C. Speakman, M.Sc., Ph.D., Senior Lecturer in Chemistry; an analysis of the difficulties inherent in the x-ray crystallographic approach to the problems of molecular structure, leading on to recent adaptations and the use of electronic computers in the study.

(7) '*Science and the Investigation of Crime*' by Professor John Glaister, J.P., M.D., D.Sc., Regius Professor of Forensic Medicine ; the rôle of the scientist as a partner in criminal investigation, with reference to the application of new active counter measures against crime.

(8) '*The Theory of Elementary Particles*' by Professor John C. Gunn, M.A., Cargill Professor of Natural Philosophy ; a general description of the present state of the theory of elementary particles and their reactions, indicating some of the outstanding problems which the theory has to accommodate.

(9) '*Educational Problems in U.S.A. and U.S.S.R.*' by Professor J. M. Robertson, M.A., Ph.D., D.Sc., F.R.S., Gardiner Professor of Chemistry ; an account of some aspects of science education and research in the United States and in Russia, with special reference to the steps which are being taken to improve science teaching in American High Schools.

(10) '*Man and Energy, with particular reference to Controlled Thermo-nuclear Research*' by Dr. S. C. Curran, M.A., D.Sc., F.R.S., Principal of the Royal College of Science and Technology, Glasgow ; a review of the problems associated with the controlled release of thermo-nuclear energy, with reference to methods by which the attack on these problems could profitably be followed.

(11) '*The Chemical Basis of Heredity*' by Professor J. N. Davidson, Gardiner Professor of Physiological Chemistry ; an explanation of the structure and function of ribonucleic acid and deoxyribonucleic acid and their significance in protein synthesis and genetics.

(12) '*An Experimental Introduction to Quantum Physics*' by Dr. K. Hecht ; a demonstration lecture showing how an experimental introduction may be given in schools to the study of quantum physics, with experiments illustrating the character of the discontinuous distribution of energy in the atomic range.

(13) '*The Work of the Conservation Corps*' by Brigadier E. F. E. Armstrong, C.B.E., Organizer of the Conservation Corps of the Council for Nature ; an illustrated review of the aims, arrangements and problems of the Conservation Corps.

(14) '*The Scalpel and the Spectroscope*' by Dr. J. M. A. Lenihan, M.Sc., Ph.D., A.M.I.E.E., F.Inst.P., Regional Physicist, Western Regional Hospital Board, Glasgow; a review of the rapidly widening knowledge required by physicians and physicists, leading to increasing specialization and the need for new co-operation between physics and medicine.

Two evening sessions were given over to lectures, supplemented by a working exhibition of laboratory work and films, by Dr. Francis L. Friedman of the Massachusetts Institute of Technology, on the subject of the work of the American Physical Sciences Study Committee. This Committee was set up in the United States to study the teaching of Physics in High Schools and to suggest such changes as might be desirable. In the lectures, Dr. Friedman outlined the structure of the P.S.S.C. and summarised the principal ideas underlying this important and highly successful venture.

The visits programme provided a wide range of activity, and among the firms who gave much appreciated hospitality were the Scottish Television Studios, Hunterston Nuclear Generating Station, Fibre Glass Ltd., J. & R. Tennent Ltd., B.P. Petroleum Refinery, British Hydrocarbon Chemicals Ltd., I.C.I. Dyestuffs Division, Babcock and Wilcox Ltd., Colville's Steel Works, John Brown & Co., Geo. Ballantyne & Son, the National Engineering Laboratory, Kelvin & Hughes Ltd., Barr & Stroud Ltd., Alex. Stephen & Sons Ltd., and the Marine Biological

Station, Millport, Isle of Cumbrae. Members were also conducted round various University Departments, and showed a keen interest in the work which was going on there.

The usual high standard was maintained in the Members' Exhibition, while the Manufacturers and Publishers provided excellent displays of new equipment and textbooks in the Chemistry Department. Following the great success of the special exhibition on Atomic and Molecular Structure and Nuclear Energy at the Southampton Meeting, a similar exhibition was put on in the new Engineering Building. In addition to some most interesting equipment shown by British firms and organisations, apparatus available from firms in the United States and in Western Germany was shown.

The programme of films in the Chemistry Department in the afternoons of Wednesday and Thurdsay provided interest and instruction, and for some a happy relaxation from the bustling round of lectures and exhibitions.

As usual there was ample scope for discussion, notably in the Business Meeting with its deliberations on new subscription rates and the pattern of the new-look for Branch activities, and in the Science and Religion Group's meeting, which was initiated by Dr. J. W. D. Smith on ' Christian Faith and Scientific Knowledge.'

There is so much to be assimilated at the Annual Meeting ; so many new ideas to pursue, so much to see and hear that is challenging and vital. But for many members the keenest pleasure lies in meeting old friends and making new ones. In all these Glasgow University was a worthy setting.

H. LOCKETT.

Editor's Note :

This report was contributed by the late Mr. H. Lockett, the devoted Annual Meetings Secretary of the Science Masters' Association. It was with deep regret that we heard of his sudden death on January 27.

D.S.I.R. Grants and Awards, 1961

The pamphlet *D.S.I.R. Research Grants*, 1961 (H.M.S.O., 1*s.*) gives the regulations governing grants for the support of research projects in the natural sciences, applied science, technology, and engineering. There is also a note on the OEEC senior visiting fellowships.

D.S.I.R. Studentships and Fellowships, 1961 (H.M.S.O., 1*s.* 9*d.*) describes the awards for postgraduate students, both for research training and for advanced study, and the research fellowships available for those who have completed the first stage in a research career. It mentions the D.S.I.R. awards that are tenable abroad, and the science studentships and fellowships financed by the North Atlantic Council which are intended to stimulate the exchange of postgraduate and more senior research workers between the NATO countries.

BOOK REVIEWS

Translation from Russian for Scientists. By C. R. Buxton and H. S. Jackson. (Blackie and Son Ltd., 1960.) [Pp. 299.] 30*s*.

It has long been realized that the study of Russian will be as important for scientists as it is for students of literature, art and drama, but there is rarely an opportunity for a scientist to start a new language at the same time as his other work. This is not the first book which has gathered together scientific readings in Russian as 'battle-practice' for the Soviet scientific journals, but it is the biggest collection I have seen and it covers the widest range of subjects : as such it is excellent. But it also claims to include enough material for the attainment of "an adequate reading knowledge of the . . . language". Granted that scientific Russian, from the linguistic point of view, is not particularly difficult, I cannot believe that this summary *by itself* will be enough to allow the reader to translate the readings given ; because it *is* only a summary, and a more detailed text-book will surely be needed by anyone starting Russian from scratch, who does not already have at any rate a good knowledge of a highly inflected language such as Latin or German. As it stands, a system of cross-referencing between the grammar sections and the exercises on them is needed, and it would help if at any rate the first part of the book were printed with stresses on the Russian text. A man learning a language, even if only to read it, pronounces the words to himself, and he must in this be consistent.

The book's shortcoming underlines the importance of introducing Russian in school curricula for as many potential university entrants as possible ; beyond ' O ' level standard, it is not a hard language whether progress is made towards Dostoievsky or the fact-packed Soviet jargon. The scientist who has done a year or two of Russian already may well find this book invaluable.

G. B. Cobbold.

Teach Yourself Atomic Physics. By J. M. Valentine. (London: English Universities Press, 1960.) [Pp. 192.] 7*s*. 6*d*.

The English Universities Press are to be congratulated on this latest addition to the ' Teach Yourself ' series. Dr. Valentine has a gift for logical arrangement and exposition of his material ; he writes clearly and simply, eschewing mathematics, but not flinching from the use of proper technical terms and symbols ; thus the reader has no sense of being ' talked down to '. Of course, no reader of *Contemporary Physics* will himself need such a book, but he should have it on his bookshelf : so convenient to quote from, or to lend to the too pressing or enquiring layman friend ! And suppose that, in the near future, the ' atomic physics ' syllabuses of the Science Masters' Association are adopted, well, while this is not a suitable *text-book*, it will be a cheap and excellent subsidiary source of information for both pupils and teacher.

The first chapter provides a concise and necessary introduction to atoms and molecules ; cathode rays and x-rays. We then get down to the main business with atomic structure (Bohr as well as Rutherford), neutrons and protons, isotopes, radioactivity and nuclear structure, followed by chapters on particle accelerators, geiger and other forms of detectors, nuclear reactions and reactors, nuclear weapons, and finally, the industrial, medical and biological uses of radioisotopes. There are several excellent and informative tables, and the line diagrams are good. One statement the reviewer would condemn : " Relativity theory shows that mass and energy are interconvertible. In other words, the mass of a particle can be wholly converted to energy . . . " ; this is surely misleading, how can mass be converted to energy when energy, in whatever form, still possesses a mass E/c^2 ? *Matter* can be converted to energy, but not mass.

This is a minor point ; in general the book excellently achieves its limited objective.

H. F. Boulind.

Analytical Techniques for Non-Linear Control Systems. By JOHN C. WEST. (London: English Universities Press Ltd., 1960.) [Pp. xii + 223.] 25*s.*

The automatic control of industrial processes (automation, as it is popularly known) is usually achieved by means of a complex servo-mechanism in which an error (i.e. the difference between the actual and the desired form of the product) is converted into an electrical signal, amplified, and then used to control the operating machinery so as to reduce the error to zero. The whole system then becomes self-regulating, and the study of self-regulating systems is of great importance not only in Physics and Engineering but in Sociology and the Biological Sciences. The behaviour of such systems can be described by a differential equation which is frequently non-linear. The time-honoured way of dealing with such equations is to linearize them, i.e., to omit the non-linear terms. All too often this means that the baby gets thrown out with the bath-water, and certain basic characteristics of the system are overlooked. The differential equations which arise in the analysis of these systems are all too often those for which mathematical techniques are lacking; and this book, the first of its kind, deals exclusively with methods for their solution. The wholly commendable spirit of this book is that any method, however approximate, which yields answers where none previously existed is worth while.

E. W. LEE.

General Science Physics. By M. NELKON. (London: Heinemann, 1960.) [Pp. 238]. 10*s.* 6*d.*

This book is intended for the final year of a course for the O-level papers in Physics-with-Chemistry or General Science, it being assumed that the previous years have been spent studying *Elementary Physics* by Abbott and Nelkon. The first part of the book therefore completes the course by dealing with more advanced topics, while the second part provides a revision course for the rest of the syllabus.

An impressive feature of the first part is the speed with which each topic is covered. One fears that the student's knowledge will be superficial—but this may be due to the extent of the syllabus rather than the author's presentation. He does however excel himself in the last chapter, by dealing with the electromagnetic spectrum, valves, television, transistors, nuclear fission and Zeta in sixteen pages. This space, as well as that devoted to glossy photographs of the manufacture of transistors or the construction of Zeta, might have been used to give a more careful treatment of simple topics.

Some detailed criticism of the text could be made, but on the other hand the book is well produced, with clear line drawings, useful summaries at the end of each chapter, and an ample supply of examples. For its particular purpose—to follow on after the previous book—it would seem reasonably sound, but pupils brought up to learn their physics in a different order would not find it much help.

W. A. MAYER

Magnetic Materials. By F. BRAILSFORD. (London: Methuen and Co. Ltd., 1960) (Methuen Monographs on Physical Subjects.) [Pp. 188.] 16*s.*

This is the Third Edition of the author's book which was first published in the Methuen Monograph series in 1948. The book has been extensively re-written and a conversion has been made to the rationalized M.K.S. system of units. Chapter headings are similar to the previous edition except that the chapters on Ferromagnetism and the Properties and Theory of Single Crystals have now been combined into a single chapter on Ferromagnetic Domains and the Properties of Single Crystals; and an additonal chapter on Ferrites has been included at the end.

Although the book has been rewritten, the reviewer felt very disappointed that the opportunity had not been taken to bring it really up to date. Apart from the chapter on Ferrites, much of the work of the last five to ten years receives scant attention in the book, and yet this surely must be one of the main reasons for preparing a new edition. This is particularly true in the case of a book the price of which has been more than doubled, so that instead of being in the 'cheap' monograph range, it is now in a price range where it will be in competition with 'paperbacks' that give very much greater value for money.

J. H. PHILLIPS.

Crystal Structure Analysis. By M. J. Buerger. (J. Wiley and Sons Ltd., 1960) [Pp. 668.] 148*s.* (New York $18.50).

Professor Buerger's reputation in the crystallographic world is such that the publication of any book by him is an eagerly awaited event. When the book is of the size and scope of his latest work it is not surprising that considerable interest should be aroused, and that it should be subjected to close and intense scrutiny. To do justice to such a work a review of some considerable length would be required, and detailed criticism would be more appropriate in a crystallographic journal than in " A Journal of Interpretation and Review " The present writer will therefore confine his remarks to the general outline and character. of *Crystal Structure Analysis.*

The first three chapters provide an introduction to the theoretical background of x-ray diffraction studies, discussing in particular the relationship with elementary diffraction theory. Chapter 4 then briefly outlines the essential steps in crystal-structure analysis, giving due emphasis to the preparatory stages and to the importance of optical and morphological examination before x-ray studies begin. The remainder of the book (nearly 600 pages) is concerned with the amplification of these steps and discussions of related topics including, for example, the measurement and correction of intensities, structure-factor calculations, Fourier theory, and some illustrations of actual structure determinations.

The volume of information contained here is enormous and in many instances practical instructions in minute detail are given. The whole flavour is that of a ' handbook ', a book to be kept close at hand during actual experiments, containing many hints and tips which could not find a place in more conventional textbooks. Some extensive and original theoretical material also finds a place.

To compile such a work is clearly a major undertaking, and it is inevitable that the particular choice of material will please some more than others. This is, in fact, a very personal book ; there are occasional gratifying glimpses of the author as a man as well as a scientist (" the correction for absorption is an exasperating one . . . . "), and it is marked by strong personal opinions. The result is that, for a general study, it is a little unbalanced; a number of subjects and approaches that are in wide use are presented in a less up-to-date form than those approved and used in Professor Buerger's own laboratories. The level of treatment also varies ; some sections begin at a very elementary level and are full of detail, whereas others plunge in at a more advanced stage.

The tendency to unbalance is to a certain extent compensated by extremely full and very valuable bibliographies at the end of each chapter. Details of papers to which specific reference has been made are given, also large numbers of papers on related topics are included and—a very useful point—titles of papers as well as authors and references are given.

For workers in large crystallographic departments the book will be an excellent source for reference on a wide range of topics, and will no doubt stimulate useful discussions. For those working alone it may prove to be an invaluable *vade-mecum.*

C. A. Taylor.

Surface Microtopography. By S. Tolansky. (Longmans, 1960.) Pp. 296. 55*s.*

The two outstanding features of this book are the extremely high standard of production and the obvious enthusiasm of the author.

The first five chapters survey the basic theory of multiple beam interferometry and draw attention to the conditions that must be fulfilled if the potentialities of the method are to be realized to the full. The author also explains some of the failures of the less skilful practitioners—the appearance of poor or unsymmetrical fringes by reflection, for example.

The rest of the book describes the application of the method to a wide variety of problems and it is here that the author demonstrates the power and beauty of his technique. Topics include growth features and etch patterns on crystal surfaces, crystal cleavages, oscillating quartz crystals, and the effect of indentations and impacts on metals; diamonds are studied in considerable detail.

The text is illustrated by over 350 excellent photographs. In fact, if one reverses the order and says that the photographs are accompanied by an interesting and informative commentary, one gives a better impression of the nature of the greater part of the book.

In conclusion, there is a useful list of papers and theses published by the author and his collaborators.

R. S. Longhurst.

High Frequency Applications of Ferrites. By J. ROBERTS. (London: English Universities. Press, 1960.) [Pp. 166.] 16*s.*

"Why is the core of a transformer laminated ?" is a well-known problem. The answer is, of course, to reduce the eddy currents which would otherwise flow and heat it, leading to inefficiency and perhaps disaster. Another approach is to suggest that the core be made of a magnetic material which is itself an electrical insulator. As a result of the discovery of ferrites (which are rather complicated oxides of transition and other metals), this has now become possible. Also as a sort of bonus the ferrites are found to have other unusual properties, and these too can be exploited in a number of previously unsuspected ways.

The present volume gives a readable, if fairly superficial, account of the general properties of ferrites and describes, with varying amounts of detail, a number of ways in which ferrites have been used. A reader who is unfamiliar with the field should find no difficulty with the text and will end by knowing that ferritse exist and that they have been used in a number of very ingenious ways. If, however, he wants to know more, perhaps because he now feels that he ought to be able to use a ferrite in some specific application, he is likely to feel frustrated, for the author has led him so far and has then abandoned him. It is true that in a small volume it is only possible to cover a limited amount of ground, but the book would have been much improved had the author given some indication of just how much there is to know about ferrites and where this may be found. K. W. H. STEVENS.

Christian Belief and Science. By R. E. D. CLARK. (London : English Universities Press Ltd., 1960.) [Pp. 160.] 12*s.* 6*d.*

In the somewhat limited field of the relationship between Christianity and Natural Science it is becoming increasingly difficult to think up a title which has not been used already by another writer. Under the circumstances we ought not to complain therefore if the title only tells us where to find a place for the book on the library shelves. In fact the book under review is a great deal more interesting than its rather prosaic title would suggest. It does not run in any of the ordinary well-worn grooves, but has something novel and important to say. For this reason it does seem to be a pity that no suggestion of the good fare within is contained in the '$mx+c$' formula of the title.

' Something novel and important to say '—something, it might be added, which a young scientist sixth-former will find stimulating and enthralling—something indeed about *science.* For this is really a book about how a research scientist succeeds in discovering anything. By the time a new discovery has found its way into the text-books, all traces of how it came to be discovered are usually covered up quite effectively. The reader of the average text-book would scarcely guess that a discoverer owed his discovery every bit as much to a flash of intuition from outside himself, as to careful planning and observation in the laboratory. The normal write-up, even of a research communication, effectively shrouds what really went on in a man's mind. We get no hint either of the patient searching in the faith that the problem is soluble, or of the sudden burst of light and joy when the key idea dawns. It is these experiences of *real* investigation that Mr. Clark's book dwells upon so fruitfully.

What is the connexion of all this with Christian Belief? Simply that Scientists and Christians use their minds in the same way—indeed there is only one way to use the mind if we are to *achieve* anything. The link between Science and the Christian life is the experience (common to them both) of exploration, discovery, faithful perseverance, and self-committal. The man who understands what science really is, will find his insights frequently illuminated in the pages of the New Testament, and in the Christian tradition the thrill of true spiritual discovery. If at times the value and significance of institutional religion tend to be depreciated it is of no great importance in a book of this character. It does not set out to be a defence of tradition, either religious or scientific. The author writes with an infectious freshness, which he succeeds in imparting to his reader. The freshness is that of discovery, and of rediscovery: it is the freshness of the Spirit ' which bloweth where it listeth '.

G. D. YARNOLD.

Science in Civil Defence

by J. McAULAY
(Scientific Advisers' Branch, The Home Office)

Summary

Some of the consequences of a nuclear attack on the United Kingdom and the problems which might arise are reviewed. Scientists would be needed to deal with many unexpected problems: their organization and duties in Civil Defence are outlined.

The effects of a nuclear attack on the United Kingdom could be very complicated and many difficult problems would arise. Hence the need in Civil Defence for scientifically trained minds capable of dealing objectively with unexpected problems.

1. Introduction

The fireball of a nuclear detonation above a critical altitude would not touch the ground. Particles of dust and debris might be sucked up into the air but they would not catch up with the rapidly ascending fireball: as the latter cooled, the vaporized contents of the weapon and its fission products would condense to minute particles that would take days or weeks to fall to the ground.

Scaling rules have been published(1, 2) for calculating the ranges at which injury to human beings, fires and structural damage of various degrees would be caused by a weapon of any given yield or power. The yield is expressed in kiloton (kt) or megaton (MT) units which represent respectively the amounts of energy released on the detonation of a thousand and a million tons of TNT.

If the fireball touched the ground, particles of earth and debris would be sucked up into the fireball and into the ' mushroom ' cloud, to act as carriers on which fission products would condense. These particles would form the radioactive ' Fall-out ' plume or pattern when they reached the ground at times which would depend upon the size of each particle, the height from which it fell and upon the mean vector wind from ground level to this height.

2. A nuclear cloud model for fall-out prediction

A paper by Schuert(3) contains calculations of the falling rates of particles of different sizes, densities and shapes from different altitudes, together with a source model of a nuclear cloud devised for predicting the paths of fall-out plumes. Using also the results of samples taken by rockets passed through the radioactive clouds at weapon trials, the parameters(4) of the cloud model have been chosen to give agreement between forecast and actual fall-out plumes for a number of nuclear detonations above 100 kilotons. Clouds from smaller weapons are more spherical and do not reach the tropopause, i.e. about 35 000 to 40 000 ft over the U.K. (cf. fig. 1).

The cloud ceases to rise about 7 to 10 min after the detonation. For weapons of more than 100 kt the height of the cloud is a logarithmic function of the weapon power and the cloud takes the form of a disk, about 25 000 ft thick irrespective of the weapon power, surmounting a stem about one-tenth the diameter of the disk. It appears that most of the activity resides in the bottom third of the disk and in the upper third of the stem, that particles greater than 1 mm diameter are confined to the stem and particles between 1 mm and 500 μ are restricted to the inner half of the cloud radius.

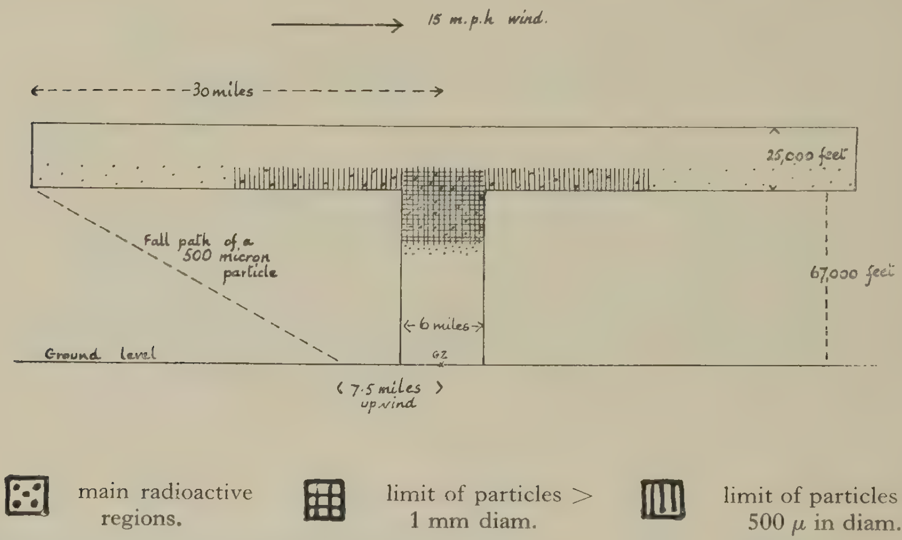

15 m.p.h wind.
30 miles
25,000 feet
Fall path of a 500 micron particle
67,000 feet
6 miles
GZ
Ground level
7·5 miles upwind
main radioactive regions.
limit of particles > 1 mm diam.
limit of particles > 500 μ in diam.

Fig. 1. Cloud model for a 10 Mt surface detonation in a 15 m.p.h. mean wind from 0 to 70 000 ft.

3. Access to damaged areas upwind for fire-fighting and rescue

The cloud model is useful in considering the extent to which radiation from fall-out could delay access by civil defence forces to the damaged area across-wind and upwind of a nuclear detonation. Figure 1 illustrates the cloud model for a 10 Mt detonation with a 15 mile per hour mean wind from 0 to 67 000 ft, i.e. to the base of the cloud. Particles 500 μ in diameter near the base of the cloud would take about an hour and a half to fall and particles of 50 μ would take over 20 h.

Assuming as an extreme case that a 500 μ particle might be found near the upwind edge of the cloud, i.e. 30 miles from its centre, it would be carried along by the wind for 22·5 miles and fall to the ground 7·5 miles upwind of the point of detonation. This would represent the maximum limit of fall-out upwind of a 10 Mt detonation with a mean wind of 15 m.p.h.

The cloud from a 1 Mt detonation would have a radius of 9·5 miles with a base at about 44 000 ft(5). It could not produce fall-out upwind of the point of detonation with a mean wind speed of 15 m.p.h. because a 500 μ particle at the upwind edge of the clould would take 1·1 h to fall and it would be carried

7 miles beyond the point of detonation and 1·5 miles beyond it in a 10 m.p.h. wind.

The records of mean wind speeds in the United Kingdom from ground level to 70 000 ft show that mean wind speeds below 10 m.p.h. occur with a frequency of 15% and those below 15 m.p.h. with a frequency of 27%. Moreover, mean wind speeds in the United Kingdom are greatest for the range 0 to about 40 000 ft.

4. CHARACTERISTICS OF FALL-OUT

Weapon trials have been done with kt detonations on desert sand and with MT detonations on coral islands. One can only guess at the nature of fall-out from large detonations on deep clay. A sample of fall-out from the American trial at Bikini on 1st March, 1954, was collected from the deck of the Japanese ship *Fukurya Maru* and sent to Shizuoka University: fourteen detailed reports of experimental investigations of the chemical and physical properties of the particles in this sample of fall-out have been published⁽⁶⁾.

The ship was 90 miles from the detonation and it was reported that " about 3 h later fine dust began to fall on the deck with a faint noise ": the fall-out lasted for about 5 h. In an examination of a sub-sample containing about 200 particles, 2% had diameters less than 40 μ and 2% greater than 400 μ. The surfaces of most of the particles were fused and contained very active black spots of 2 to 3 μ diameter.

The most interesting observation from electron diffraction and from X-ray analysis was that the particles consisted of calcite with no apparent trace of aragonite. Since Bikini coral is 100% aragonite the particles could not have resulted from the disruption of unchanged coral only but must have been decomposed and possibly also vaporised in the fireball.

5. THE DECAY OF RADIOACTIVE FALL-OUT

Each radioactive isotope follows a characteristic exponential decay law: fall-out consists of many different isotopes and has an average decay rate such that the activity at any time t h after the detonation can be represented approximately by $R^1 t^{-1 \cdot 2}$ where R^1 is the activity at 1 h.

Estimation of the rate of decay of radioactive fall-out is complicated by a number of factors. Firstly, a considerable fraction of the activity in the early cloud escapes in the form of gaseous fission products; secondly, neutron capture by U^{238} present in the weapon produces the additional radioactive products U^{239}, Np^{239}, Pu^{239} and U^{237} (n, $2n$ reaction); thirdly, the neutron-induced activity in the earth and debris, e.g. Na^{24} and Mn^{56}, may make a considerable contribution to the dose-rate on the ground for several days after the detonation. Finally, there are the unpredictable effects of weathering by wind and rain.

The decay of fall-out is discussed in detail by Triffet and by Knapp⁽⁷⁾, who refer to Miller's new data on U–235 fission products: these indicate that the initial total γ activity per kiloton of yield may be greater by a factor of 2·5 than has been hitherto assumed. Nevertheless, the $t^{-1 \cdot 2}$ decay formula, under practical conditions of use at Civil Defence operational levels, remains adequate for calculating changes in dose-rates at any point in a fall-out field for the first three months after the detonation. But unexpected variations might occur because of the number of factors involved and, for vital operations such as the cleaning of a heavily contaminated area or the removal of people from that area, plans should be based on dose-rate measurements made not more than about 12 h before the operation is started.

6. HAZARDS FROM FALL-OUT

A dose of radiation absorbed in the body is usually expressed in 'rad' units (100 ergs of energy in 1 gm of tissue) but for Civil Defence purposes the terms rad and röntgen are used indiscriminately: the intensity or dose-rate in a radiation field is expressed in röntgens per hour (r.p.h.). The curie unit is a measure of quantity of radioactive matter present in any particular place.

Fall-out could present three separate hazards to human life, (*a*) whole body effect of the environmental γ radiation which would be the predominant hazard in war, (*b*) skin burns caused by β-emitting isotopes in fall-out particles on the skin and clothing and (*c*) long-term effects, e.g. leukaemia or possible tumour developments many years later as a result of the whole body γ dose plus inhaled or ingested α-or β-emitters of relatively long radiological life which can accumulate in specific body organs and remain there for long periods.

The environmental γ radiation would come from large quantities of fall-out at distances out to several hundred feet. Anyone exposed to fall-out in the open would suffer much greater permanent injury from the environmental γ radiation than from contact with fall-out particles on the skin and clothing. Beta radiation is attenuated by ordinary clothing, nevertheless prolonged contact with heavy (and visible) fall-out in its early stages could result in skin burns. Shelter against γ radiation is the first essential; thereafter, the skin hazard can be removed by taking off the outer clothing and washing the exposed skin.

The internal hazard is difficult to assess but the following factors indicate that it would be a minor one compared with the other hazards of a nuclear war. Firstly, particles must be about 1 to 5 μ in diameter for significant retention to occur in the lungs. Particles of this size range would take so long to fall, and would be carried by the wind and spread so far, that when they reached the ground the inhalation hazard would be reduced to relative insignificance. Secondly, larger particles, when ingested, pass through the gastro-intestinal tract in about 30 h: under extreme conditions β-emission might cause inflammation of the lining of the gut. Thirdly, only a small fraction of the activity in the solid particles within the measurable fall-out pattern could be leached out by body fluids and of this soluble portion less than a quarter would be retained in the body: the fraction would be much smaller if ingestion occurred indirectly through the food chain. A few special problems would remain, e.g. the hazard to infants of radioactive iodine in milk because the iodine is concentrated in a very small thyroid gland, the ingestion of α emitters which accumulate and remain in the skeleton, such as the isotopes of U and Pu mentioned in section 5 and the problems of Sr-90, C-14 and Cs-137 in human environments throughout the world long after a nuclear war.

7. PROTECTION FROM FALL-OUT

This is based on three principles, viz. material *shielding* to attenuate the radiation especially during the early period of rapid *decay*, combined with maximum *distance* from the radioactive fall-out to take advantage of the inverse square law, i.e. the proportionality between dose and inverse square of the distance. The application of these principles, with some simplifying assumptions, has led to a quick method for calculating the protection afforded in any selected refuge room in a house or building from the dimensions and shape of the building and the weight of the material used in its construction. The radiation which penetrates into the refuge room is split up into five separate and additive contributions from fall-out on the roof and beyond each of the four external walls.

The Protective Factor is defined as the inverse of the ratio of the dose-rate inside the refuge to that 3 ft above a large open space covered by a fall-out deposit of the same specific activity per unit area. A memorandum on this method for calculating the protective factors of buildings has been supplied by the Home Office to all local authorities and major industrial concerns: an abbreviation of the method for application to typical British homes has been published(8).

8. THE PROBLEM OF DECONTAMINATION

Water-droplet fall-out resulting from a nuclear detonation on an extensive water surface must be distinguished from solid particle fall-out from a detonation on land. Fission products in colloidal solution penetrate or adhere so firmly to most surfaces that removal of the radioactivity requires abrasion of the surface or chemical treatment with acids, detergents or chelating compounds which could have only limited application in Civil Defence. Fission products are either fused into solid particles or they adhere to them so strongly that, with a few exceptions, only a little of the activity is transferred and fixed to the surface on which the fall-out is deposited, even in the presence of water.

It is thought that solid particle fall-out is much more likely to occur after a nuclear attack on the United Kingdom than the more localised fall-out from detonations on open water: the major problem would be, therefore, the mechanical removal of the radioactive particles to a place where they could be left to decay or burial to a safe depth, e.g. by deep ploughing or covering with uncontaminated earth.

Field trials have been done in the U.S.A. on the decontamination of typical roofs and road surfaces using commercial mechanized equipment for road sweeping and flushing and also by hand brushing and firehosing: bulldozing, grading, scraping, ploughing and filling equipment have been tried on unpaved surfaces. The U.S. Services Radiological Recovery Manual TP–PL–13 referred to in the Congress Hearings (9) contains a description of the various procedures used and the results of effectiveness measurements. References to other studies including two carried out on behalf of the Office of Civil and Defence Mobilization (10, 11) are included in the Appendix to the June 1959 Congress Hearings.

9. THE ORGANIZATION AND DUTIES OF SCIENTISTS IN CIVIL DEFENCE

The Civil Defence Corps includes among its members about 1500 Scientific Intelligence Officers with university degrees or equivalent professional qualifications. Their initial training is in general Civil Defence organization and duties: this is followed by a second course, in the local University, on the scientific aspects of nuclear, biological and chemical (ABC) defence. The second course is arranged by Regional Scientific Advisers who are leading scientists in each region. In a third, more practical course, they study and seek solutions to the problems which could arise after a nuclear attack. But the most valuable service which S.I.O.s can render in peacetime is to take part in regional or local exercises and to provide constructive criticism essential to the development of a realistic civil defence doctrine. The primary task of S.I.O.s is to advise on all scientific and technical aspects of ABC defence at every level of operational control and to make sure that each Controller is fully cognisant of the risks and chances of success associated with any proposed course of action.

10. THE WARNING AND MONITORING ORGANIZATION AND THE ROYAL OBSERVER CORPS

The scale of nuclear attack might be such that reinforcement of fire-fighting and life-saving operations would have to be directed from a regional or even a national level. Much would be done locally, of course, until the radiological instruments showed that it was imperative to seek refuge. Major Civil Defence resources, however, could not be used effectively in the absence of information about the overall pattern of the attack, the location of each burst, its power, how much fall-out could be expected, the likely paths of all the fall-out plumes and the possibility of finding routes and areas of sufficiently low radiation intensity for the transport and deployment of resources.

Early information about the nuclear attack would come from the network of underground R.O.C. posts equipped with instruments for the purpose. Each post would report to the appropriate underground Sector Operations Centre where the data would be processed and plotted on maps by Sector Scientists with similar qualifications to the S.I.O.s.

11. FALL-OUT PREDICTION

Meteorologists at each Centre would prepare periodic forecasts of mean vector winds from ground level up to various heights (80 000 ft and higher if possible). These forecasts would enable the Sector Scientists to cut out cardboard templates which could be placed on a map to show the predicted limits of the fall-out areas from nuclear bursts.

A template is made by marking the N–S line on the cardboard and drawing a circle with its centre as the point of burst (G.Z.) and its radius to represent the maximum likely upwind distance of the fall-out: the latter is estimated from a chart containing a family of curves relating this parameter to bomb power and to the lowest of the mean vector wind speeds. Lines representing the extreme wind bearings are drawn tangential to the circle (fig. 2). In making the template it is wise to add margins for possible errors in wind speeds and bearings.

The template shows the area in which fall-out might occur in the first hour after stabilization of the radioactive cloud. It can be extended on the map downwind to show the predicted positions of the fall-out front at subsequent hourly intervals. By turning the template through 180° it can be used to mark on the map the boundary within which any ground burst bomb could cause a threat of fall-out to any of the Sector territory.

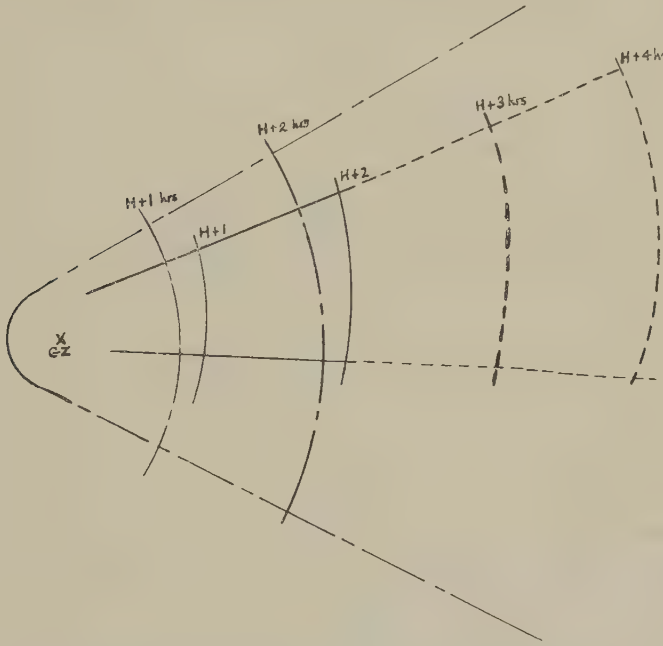


Fig. 2. Fall-out forecasts and predictions.

— — — — — forecast limit of fall-out for preliminary warning and estimated arrival times.

———— fall-out reports received and actual arrival times.

- - - - - - - improved prediction for second warning: fall-out will arrive at times estimated.

12. Fall-out warnings

An immediate black warning (fall-out imminent) would be issued to all districts within the sector scientist's template prediction. Later, as reports were received that fall-out was occurring at some posts and not at adjacent ones, the boundaries of the actual fall-out plumes could be plotted and extended to show more accurately the predicted paths of the fall-out plumes. This would permit grey warnings (fall-out expected in 1 h) and black warnings to be issued in advance of the passage of the fall-out front over the countryside (fig. 2).

13. Fall-out contours

As soon as R.O.C. post readings indicated that the local dose-rate had reached a maximum, the Sector Scientist could calculate and plot on the map the dose-rate as at some agreed reference time, usually seven hours after the detonation [H +7 h]. He could then draw contours joining points of equal dose-rate at this reference time and, when the decay function was known, calculate the dose-rate at any desired location within the fall-out plume at any future time. Plumes from different bombs could overlap and cause the dose-rate at some R.O.C. posts to rise again. The dose-rates at each post would be plotted at 5 min intervals and from these plots and the fall-out predictions of adjacent upwind Sectors, the Sector Scientist could estimate the separate dose-rate contributions from each plume. This would be necessary for estimating dose-rates at some future time since the rates of the individual contributions could vary considerably with time unless the detonations all occurred within a period of 1 or 2 h.

14. OPERATIONAL INFORMATION SUPPLIED TO REGIONAL SCIENTIFIC ADVISERS AND TO S.I.O.S

For civil defence purposes the United Kingdom is divided into eleven Regions (England and Wales), three Zones (Scotland) and Northern Ireland If necessary each could function as an autonomous unit.

The Regional Scientific Adviser in each region or zone would get from adjacent Sector Operations Centres information on the location, time and power of each detonation. He would also get data on the time of arrival and time of completion of fall-out at each R.O.C. post within and adjacent to his region, the calculated dose-rate over the post at the agreed reference time and the decay exponent measured from the dose-rate time record for each post.

As soon as it was received at the Regional Headquarters this information would also be passed down the Civil Defence chain to the S.I.O.s at the various operational levels. At the lower levels supplementary information from local civil defence sources or from reconnaissance parties would enable more detailed plots to be made on larger scale maps and the S.I.O. at each level would build up, as quickly and completely as possible a plot of the radiological situation within and adjacent to his own territory.

Thus wherever it was possible to conduct life-saving and other Civil Defence operations on a major scale there would be a scientist both ready and in a position to give the best advice available to those responsible for controlling the operation.

REFERENCES

(1) *The Effects of Nuclear Weapons*, June, 1957, U.S. Atomic Energy Commission (available from H.M.S.O.).

(2) *Manual of Civil Defence*, 1959, Volume I, Pamphlet No. 1, *Nuclear Weapons* (2nd Edition), H.M.S.O.

(3) Hearings before a Special Sub-Committee on Radiation of the Joint Committee on Atomic Energy of the Congress of the U.S.A., Sessions May and June, 1957, *The nature of radioactive fall-out and its effects on man*, part I, 280–307.

(4) See ref. (2), para. 8.24.

(5) See ref. (2), Appendix 2 for scaling rules.

(6) *The Bulletin of the Institute for Chemical Research of Kyoto University*, November, 1954, Supplementary Issue.

(7) U.S. Congress Hearings. Session June, 1959, *Biological and environmental effects of nuclear war*, 61–126 and 189–229 (general discussion).

(8) See ref. (2), chapter ix.

(9) U.S. Congress Hearings, 1957 Sessions, Part I, p. 265.

(10) Stanford Research Institute, November, 1958, *Systems analysis of radiological defense*, SRI Project No. IU 2324.

(11) Technical Operations Inc., July, 1958, *Radiological defense planning guide*, TOI–58–26.

The Author:

Dr. McAulay is a Principal Scientific Officer in the Scientific Advisers' Branch of the Home Office. After some ten years on industrial research he became a defence scientist in 1936, during the war as a ' special ' fuel technologist and head of the Materials Section of the Admiralty Scientific Headquarters, and subsequently in the Ministry of Defence and Ministry of Supply on toxicological warfare until the thermonuclear threat became predominant.

Satellite Orbits and their Geophysical Implications

by D. G. KING-HELE
Royal Aircraft Establishment, Farnborough

SUMMARY

Small changes in the orbits of satellites near to the earth have revealed new information about the earth's gravitational field and atmosphere. This article explains the main properties of satellite orbits and describes the discoveries which have been made with their help in the following branches of geophysics: the figure of the earth and the gravitational field; the density of the upper atmosphere and its variations, which have proved to be linked with solar activity; and, finally, other properties of the atmosphere, such as temperature and winds.

1. INTRODUCTION

The advent of artificial earth satellites has sharply stimulated progress in many branches of geophysics. Some of the discoveries—for example the zones of radiation about the earth—have depended on measurements by instruments aboard the satellites; but a great deal of information, less spectacular but just as valuable, has come indirectly, by studying the small changes in the orbits of satellites. Every satellite, whether instrumented or not, inevitably feels the effects of air resistance and of irregularities in the earth's gravitational field. By analysing the consequent small changes in the orbit, the properties of the upper atmosphere and the gravitational field have been elucidated in far more detail than was previously possible. The striking advances in our knowledge of these two subjects brought about by satellites have also, inevitably, provoked some stimulating new hypotheses.

In the last 300 years many great mathematicians, from Newton and Lagrange to Poincaré and Einstein, have made important contributions to the theory of celestial mechanics, and it might be supposed that a subject with such an illustrious history would have included the theory of earth-satellite orbits somewhere within its extensive boundaries. In fact, however, the orbit of a close earth-satellite does not conform to the usual assumptions of celestial mechanics, and its theory has had to be developed *ab initio* in the last few years for three reasons. First, it is much closer to the parent body than any known natural satellite, so that, in developing the theory, the departures of the earth's gravitational field from exact radial symmetry have to be analysed in far more detail than is usual: the main effect of the earth's oblateness, for example, is greater for a near satellite than for the moon by a factor of 10^6; and for some of the higher-order effects this factor increases to over 10^{12}. Second, the orbit of an artificial satellite is often inclined at a large angle to the equator, and the small-angle approximations which abound in celestial mechanics, e.g. in lunar theory, become invalid. Third, the orbit is seriously affected by the earth's atmosphere, and, in contrast with the quasi-permanent orbits of astronomy, usually comes to a fiery end within a few years.

In this article the main features of earth-satellite orbits revealed by theory are described first, in section 2. The new information about the earth's gravitational field is discussed in section 3, and sections 4–7 deal with various properties of the upper atmosphere, its average density, the variations in density, temperature and winds.

2. SATELLITE ORBITS

If the earth were spherically symmetrical and had no atmosphere, its gravitational attraction would be the same as that of a point mass M at its centre; and, under Newton's law, the attraction per unit mass at a distance r from its centre would be GM/r^2, where G is the universal constant of gravitation ($GM = 398600$ km^3/sec^2). The simplest orbit in this field of force is a circle, described at a constant velocity v_c, say. Since the force responsible for the inward radial acceleration, v_c^2/r, is produced by, and equal to, the gravitational attraction, it follows that

$$v_c^2 = GM/r.$$

Hence the speed for a circular orbit at a height of 500 km above the earth's surface ($r = 6870$ km) is 7·62 km/sec; at a height of 1000 km the required speed falls to 7·35 km/sec.

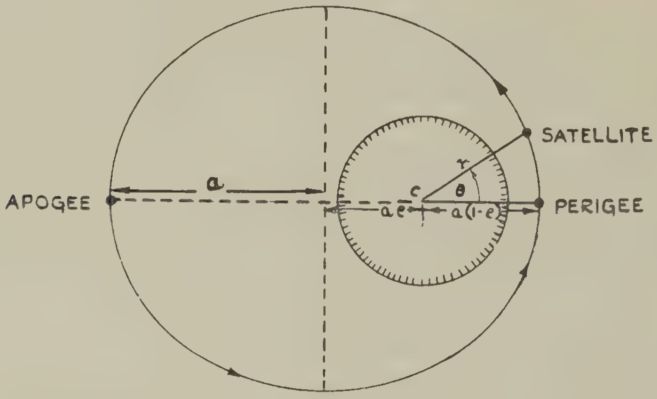


Fig. 1. Diagram of satellite orbit about the earth.

The time T taken by the satellite to complete one circuit round the earth, the *orbital period* as it is called, is obtained on dividing the circumference of the circular orbit, $2\pi r$, by the velocity, to give

$$T = 2\pi r^{3/2}/\sqrt{(GM)},$$

which may conveniently be written in numerical form as

$$T = 84{\cdot}3(r/\mathbf{R})^{3/2} \text{ minutes},$$

where $\mathbf{R}$ is the earth's mean radius (taken as 6370 km). The period of revolution in a circular orbit is 94·4 min for a height of 500 km and 104·9 min for a height of 1000 km.

In practice it is not possible to achieve exactly the correct speed for a circular orbit. If the satellite is projected into orbit, horizontally, at rather too high a speed, it swings out to a greater distance on the opposite side of the earth, and the orbit is an ellipse, with the point of minimum height, or *perigee*, at the point of entry into orbit, and the maximum height, or *apogee*, at the opposite side of the earth (fig. 1). The size and shape of the ellipse are conveniently expressed by its *semi major axis*, a, (fig. 1) and *eccentricity*, e. The eccentricity of the orbit can most usefully be defined as the difference between maximum and minimum height, divided by the major axis, $2a$; for an elliptic orbit e may take any value between 0 and 1. From this definition it follows that the distance of the satellite from the earth's centre varies between $a(1-e)$ at perigee, and $a(1+e)$ at apogee; and the polar equation of the ellipse takes the well-known form

$$r = \frac{a(1-e^2)}{1+e\cos\theta},$$

where θ is the angle SCP subtended at the earth's centre C by the satellite S and the perigee point P, as shown in fig. 1, so that the satellite passes perigee when $\theta=0$ and apogee when $\theta=180°$. If $e=0$ the ellipse reduces to a circle of radius a.

By a happy chance the orbital period T of a satellite travelling in an elliptic path depends only on a, and not on e, being given by the same equation as for a circular orbit, with r replaced by a. The velocity v of the satellite in an elliptic orbit is readily found from the energy equation. The sum of the kinetic energy per unit mass, $v^2/2$, and the potential energy, $-GM/r$, must be a constant, whose value, $-GM/2a$, is determined by the need for the equation to reduce to the same form as that for a circular orbit when $r=a$. So

$$v^2 = GM\left(\frac{2}{r}-\frac{1}{a}\right).$$

This equation shows that for any given initial distance r from the earth's centre, the velocity required to escape from the earth, i.e. to achieve an orbit with an infinite value of a, is $\surd 2$ times the circular orbital velocity of $\surd(GM/r)$. Thus, if the initial height is 500 km, an initial velocity between 7·62 km/sec and 10·78 km/sec will give an elliptic orbit; any greater velocity leads to a hyperbolic orbit and escape from the earth.

We have so far assumed that the earth is spherically symmetrical and has no atmosphere: in these conditions the orbital ellipse remains fixed in size and shape, and lies in a plane, which passes through the earth's centre and has a direction fixed in space (if we ignore small effects, such as those due to the moon and sun). When the earth's oblateness and the drag of the atmosphere are taken into account, however, this simple picture has to be modified.

It might be thought at first sight that the gravitational attraction of the earth's equatorial bulge would tend to reduce the angle at which the orbital plane cuts the equator, the *orbital inclination* as it is called. But the satellite in its orbit is not unlike a point on a spinning top, and, just as a top precesses about a vertical axis instead of falling to the floor, so the plane of the satellite orbit rotates, or precesses, about the earth's axis, while still remaining at a fixed inclination to the equator. The orbital plane rotates in the direction opposite to the satellite's motion, so that the point where an eastbound satellite crosses the equator moves steadily westward, relative to the stars, as shown in fig. 2. The rate of rotation is about

$$9{\cdot}97\left(\frac{R}{a}\right)^{1\cdot5}\left(\frac{R}{p}\right)^{2}\cos i \text{ degrees per day},$$

where $p=a(1-e^2)$, i is the orbital inclination and R is the earth's equatorial radius. The rate of rotation therefore lies between zero for a polar orbit and about 8 degrees per day for a near-equatorial orbit close to the earth.

The gravitational attraction of the earth's equatorial bulge also causes a rotation of the orbital ellipse in its own plane. The major axis rotates at a rate, measured in the same direction as the satellite is moving, of

$$4{\cdot}98\left(\frac{R}{a}\right)^{1.5}\left(\frac{R}{p}\right)^{2}\ (5\cos^2 i-1)\ \text{degrees per day.}$$

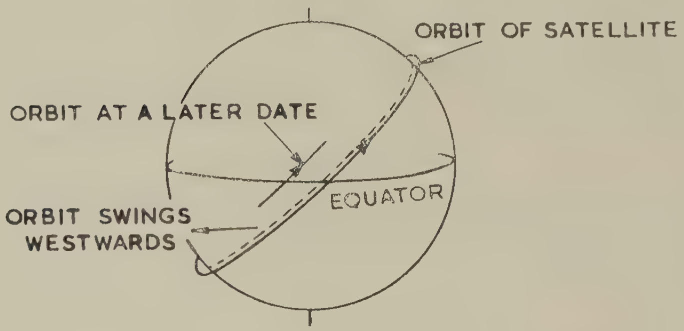


Fig. 2. Precession of orbital plane of a satellite due to the earth's equatorial bulge.

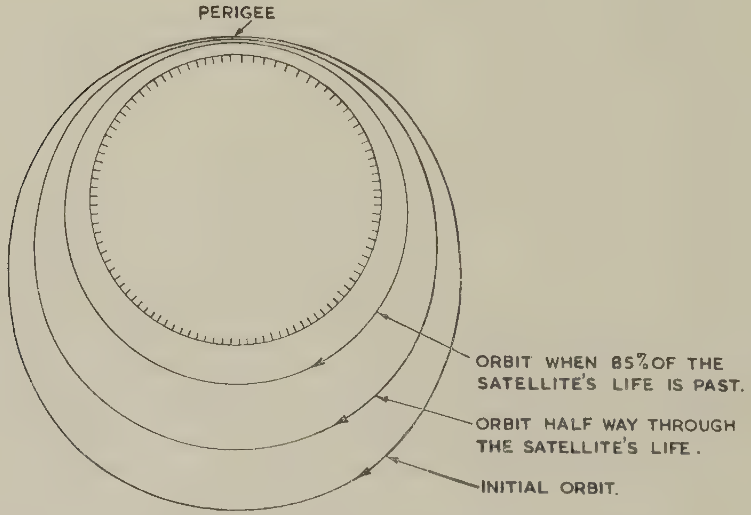

PERIGEE
ORBIT WHEN 85% OF THE SATELLITE'S LIFE IS PAST.
ORBIT HALF WAY THROUGH THE SATELLITE'S LIFE.
INITIAL ORBIT.

Fig. 3. Contraction of satellite orbit under the action of air drag.

The major axis may therefore rotate as fast as 15 degrees per day forward, i.e. anti-clockwise in fig. 1, for a near-equatorial orbit, or 4 degrees per day backwards for a polar orbit. The rotation is zero at an orbital inclination of 63·4°, where $\cos i = 1/\sqrt{5}$, and was very slow for Sputniks 1–3, whose orbital inclinations were near 65°.

Air drag affects the orbit in a quite different manner. Since air density decreases rapidly as height increases, the air drag encountered by a satellite during the small section of the orbit where it comes nearest to the earth is much greater than at any other point in the orbit. So, every time the satellite passes perigee it suffers a slight retardation, with the result that it does not swing out to quite such a great distance on the other side of the earth as on its previous revolution. The effect of air drag is therefore to reduce the height at apogee while scarcely altering the height at perigee: the orbit contracts and becomes more nearly circular, as shown in fig. 3; then, after a short time in a near-circular orbit, the satellite makes its final plunge through the atmosphere. Since the orbit gradually sinks lower into the atmosphere, its rate of contraction becomes more rapid as time goes on. It is found that, if the initial eccentricity is less than 0·2, the square of the eccentricity varies almost linearly with time t, so that the variation of eccentricity e during the lifetime t_L of the satellite is given by

$$e = e_0\sqrt{\left(1-\frac{t}{t_L}\right)},$$

where e_0 denotes the initial value. Since the perigee distance, $a(1-e)$, remains nearly constant, and the period T is proportional to $a^{3/2}$, T is, very nearly, proportional to $(1-e)^{-3/2}$. Thus, to a first approximation, the variation of orbital period with time is given by

$$\frac{T}{T_0} = 1-\frac{3}{2}(e_0-e) = 1-\frac{3}{2}e_0\left\{1-\sqrt{\left(1-\frac{t}{t_L}\right)}\right\},$$

again assuming that $e_0 < 0{\cdot}2$. The above equations are derived on the assumptions that the properties of the atmosphere do not vary from day to day, and the effective cross-sectional area of the satellite remains constant. Both these assumptions are liable to be in error, but the formulae are nevertheless of useful accuracy.

3. THE EARTH'S GRAVITATIONAL FIELD

We have already seen how the earth's equatorial bulge causes the plane of a satellite's orbit to rotate about the earth's axis. The greater the equatorial bulge, the greater would be the rate of rotation; so, by measuring the observed rate of rotation, a process which can be performed with excellent accuracy (about 1 part in 10 000, for a typical satellite), we can determine the extent of the equatorial bulge. This is the basis of the method, though the above description is, needless to say, somewhat over-simplified.

The gravitational potential outside a nearly spherical body such as the earth is conveniently expressed as the sum of a number of components. The first and by far the largest component is that due to an exact sphere, $-GM/r$; the second component represents the main effect of the equatorial bulge; the third takes account of asymmetry about the equator; and so on. Mathematically, the potential, assuming it to be independent of longitude, may be written in terms of Legendre polynomials P_n as

$$U = -\frac{GM}{r}\left\{1+\sum_{n=2}^{\infty} J_n\left(\frac{R}{r}\right)^n P_n(\cos\lambda)\right\},$$

where λ is the co-latitude, and the J_n are constants to be determined. J_1 is zero if the equatorial plane passes through the earth's centre of mass. Before the launching of satellites the only one of the J_n whose value was known was J_2, which had been determined from gravity surveys, the moon's motion and triangulations, with an accuracy of about 1 part in 300. It was usual to take also a conventional value of J_4, chosen to make the earth's meridional section an exact ellipse. The accepted pre-satellite values were $J_2 = 1091 \times 10^{-6}$, $J_4 = -2{\cdot}4 \times 10^{-6}$, though it was usual to express these in terms of the coefficients $J = 1{\cdot}5\, J_2 = 1637 \times 10^{-6}$ and $D = -35 J_4/8 = 10{\cdot}6 \times 10^{-6}$, since the J_n notation is a new development, made necessary by the satellite discoveries.

The rate of rotation, X, of the orbital plane of a satellite moving in the gravitational field specified by the potential U in the equation above can be expressed in terms of the J_n by means of orbital theory. The appropriate expression for X, carried as far as terms in J_4, is

$$X = \sqrt{\left(\frac{GM}{a^3}\right)}\left(\frac{R}{p}\right)^2 \cos i \left\{\frac{3}{2}J_2 + \frac{9}{4}J_2^2\left(\frac{R}{p}\right)^2\left(\frac{19}{12}\sin^2 i - 1\right) + \frac{3}{8}J_3 e\left(\frac{R}{p}\right)\sin w \operatorname{cosec} i\,(15\sin^2 i - 4) + \frac{15}{16}J_4\left(\frac{R}{p}\right)^2\left(7\sin^2 i - 4\right)\right\},$$

where w is the argument of perigee, the angle from the satellite's northward crossing of the equator to perigee, measured along the orbit. (The simpler formula quoted in section 2 represents the first term on the right-hand side of the above equation.)

During the life of a typical satellite p changes by only about 10% and i changes scarcely at all; so the factors multiplying J_2, J_2^2 and J_4 in the above equation do not vary much during a satellite's life; and the same applies to the factors multiplying $J_6, J_8 \ldots$ The J_3 term, however, contains the factor $e \sin w$, which is small, since e is usually about 0·1, and variable, since w varies widely and e decreases to zero at the end of a satellite's life. Thus it usually turns out that the J_3 term is negligible; and the same applies to $J_5, J_7 \ldots$ The equation is thus essentially a means of determining the even J_n. One satellite will normally give only one equation between the even J_n; two satellites of markedly different orbital inclinations will give two equations, from which J_2 and J_4 can be determined, if we assume J_6, $J_8 \ldots$ are zero; three satellites of different inclinations will give J_2, J_4, J_6, if $J_8, J_{10} \ldots$ are assumed zero; and so on. Since the J_n soon become small as n increases, the process converges rapidly.

Two satellites, Sputnik 2 and Vanguard 1, have been particularly suitable for this purpose, since both were well observed and their orbital inclinations were substantially different, 65° and 34° respectively. The values of J_2 and J_4 obtained from these two satellites, on the assumption that $J_6, J_8 \ldots$ were zero, were

$$J_2 = (1083{\cdot}1 \pm 0{\cdot}2) \times 10^{-6};\; J_4 = (-1{\cdot}4 \pm 0{\cdot}2) \times 10^{-6}.$$

The errors quoted are standard deviations arising from observational errors: there may possibly be larger errors from the neglect of J_6, J_8, etc.

An approximate value of J_6 has been obtained with the aid of a third satellite, Explorer 4, which had a most satisfactory inclination, 50°, but was not very well observed. The values of J_2, J_4 and J_6 found in this way were:

$$J_2 = (1083{\cdot}0 \pm 0{\cdot}2) \times 10^{-6};$$
$$J_4 = (-1{\cdot}3 \pm 0{\cdot}2) \times 10^{-6};$$
$$J_6 = (-0{\cdot}1 \pm 1{\cdot}5) \times 10^{-6}.$$

In deriving these values, due account was taken of (1) lunar and solar perturbations, (2) the rotation of the atmosphere, which slightly alters the rate of rotation of the orbital plane by creating a side-force on the satellite, and (3) the effect of J_3, whose value has been found by other methods, as we shall see shortly.

An alternative method for deriving J_2 and J_4 is to use the rate of rotation of the orbital plane and of the major axis for a single satellite. The disadvantage of this method is that the perigee point, and hence the rate of rotation of the major axis, is difficult to determine exactly. The method has been successfully applied to Vanguard 1, however, and the results obtained are:

$$J_2 = (1082{\cdot}5 \pm 0{\cdot}1) \times 10^{-6};$$
$$J_4 = (-1{\cdot}7 \pm 0{\cdot}1) \times 10^{-6}.$$

The errors quoted are those arising from the observations; again, errors due to neglecting J_6, etc., may be larger. These values are in satisfactory agreement with those already quoted.

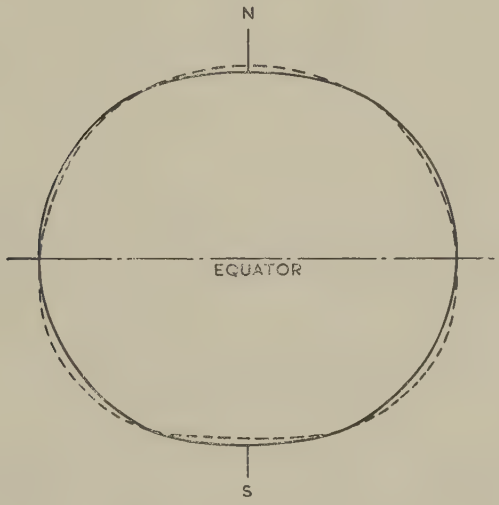


Fig. 4. The earth is not symmetrical about the equator (unbroken line), but has a very slight tendency towards a pear shape, as shown by the broken line. (Not to scale.)

The coefficients J_3 and J_5, which express any asymmetry between the southern and northern hemispheres, cause small periodic variations in the distance of perigee from the earth's centre, in the orbital inclination, and in some other orbital parameters. The amplitude of the oscillation in perigee distance caused by J_3 is $(J_3 R \sin i)/2J_2$, and it has the same period as w. Such an oscillation, with an amplitude of 3 miles, was observed with Vanguard 1, from which it was deduced that $J_3 = -2{\cdot}2 \times 10^{-6}$. Subsequent, more detailed, analysis along the same lines has indicated that

$$J_3 = (-2{\cdot}4 \pm 0{\cdot}3) \times 10^{-6}$$
$$J_5 = (-0{\cdot}05 \pm 0{\cdot}1) \times 10^{-6}.$$

So the first two years of research with satellites have improved our knowledge of the earth's gravitational field to a surprising degree. Previously, only J_2 was known, with an accuracy of 1 in 300. Now J_2 is known with an accuracy which may be as good as 1 in 5000, and values of J_3, J_4, J_5 and J_6 have been found, most of them probably with an accuracy of better than $0{\cdot}5 \times 10^{-6}$. As more satellites are launched, more values of the J_n should be established.

Our improved knowledge of the earth's gravitational field carries with it certain implications about the earth's shape, that is, the form of the earth's sea-level surface, the *geoid* as it is called. If the geoid is assumed to be exactly spheroidal, the earth's *flattening* f, defined as the difference between the equatorial and polar diameters divided by the equatorial diameter, can be simply expressed in terms of J_2. The pre-satellite value of J_2 gave $1/f = 297{\cdot}1$; the value $J_2 = (1083{\cdot}0 \pm 0{\cdot}2) \times 10^{-6}$ gives $1/f = 298{\cdot}21 \pm 0{\cdot}03$. This means that the difference between the equatorial and polar diameters is less by about 160 metres than was previously thought, 42·78 km instead of 42·94 km.

The discovery that J_3 has a non-zero value means that the earth is not symmetrical about the equator: it has a slight tendency towards a pear-shape with the stem at the north, as shown in fig. 4. Sea level at the north pole is about 30 metres further from the equator than sea level at the south pole. The presence of such asymmetry implies that there are considerable stresses inside the earth, a fact which may well influence future theories of the earth's interior.

4. THE AVERAGE DENSITY OF THE UPPER ATMOSPHERE

The rate at which a satellite orbit contracts under the influence of air drag depends on the retardation which it suffers as it passes the neighbourhood of perigee, as explained in section 2. This retardation in turn depends mainly on two factors:

(1) The density of the air at heights near that of perigee: the denser the air, the greater the retardation.

(2) The mass/area ratio of the satellite m/S, that is, its mass divided by the mean effective cross-sectional area perpendicular to its direction of motion: the greater m/S, the less the retardation.

So, if the rate at which the orbit contracts is measured, and the value of m/S is known, an estimate can be made of the air density at heights near that of perigee, i.e. at heights between 200 and 700 km.

In practice there is no difficulty in determining the rate of contraction of the orbit from observations. The orbital period can be found with excellent accuracy, about 1 part in 20 000, even from the simplest observations, which need nothing more than good eyesight, a stop-watch and a knowledge of the stars. The orbital period is, as we have already seen, related to the major axis of the orbit, and the rate of decrease of period therefore indicates the rate of contraction of the orbit. The rate of decrease of period averaged over a time-interval of, say, a month can therefore be used to give the average density during this time.

The value of the mass/area ratio is not so easily found, except for spherical satellites like Sputnik 1 and Vanguards 1 and 2. Most satellites, however, have been shaped more like long, narrow cylinders and have been free to rotate in whatever manner they pleased. Almost always they have chosen to rotate about an axis of maximum moment of inertia, i.e. an axis perpendicular to their length. The direction of this axis in space is generally unknown: it may be in line with the direction of motion, so that the satellite spins like a propeller; or at right angles to the direction of motion, so that the satellite tumbles end over end; or the axis may point in any direction between these extremes. Fortunately, whatever the direction of the axis, the effective cross-sectional area does not differ from its mean value by more than 20%, and so, even for a long thin satellite, the effective cross-section can be estimated with a standard error of about 10%.

Once the mass/area ratio, m/S, is known, the air density at perigee height, ρ_P, may be found in terms of the rate of decrease of period, dT/dt, from the formula

$$\rho_P = -\frac{dT}{dt}\frac{FSC_D}{m}\sqrt{\left(\frac{2e}{\pi a H}\right)}\left\{1-2e-\frac{H}{8ae}+\text{terms of order}\left(e^2, \frac{H^2}{a^2e^2}\right)\right\},$$

where F is a factor allowing for rotation of the atmosphere (lying between 0·9 and 1 for an eastbound satellite), and C_D is a drag coefficient, whose value usually lies between 2·1 and 2·3. In deriving this formula it is assumed that $0{\cdot}02 < e < 0{\cdot}2$ and that the density ρ varies exponentially with height y above the earth, being given by $\rho = const \times \exp(-y/H)$, where H is assumed constant. H is therefore a measure of the gradient of density: its value is not known accurately at heights above 200 km, thus limiting the accuracy of the equation for ρ_P.

This difficulty can be avoided, however, by evaluating the density at a height somewhat above perigee. Let H^* be the best estimate of H. Then the density, ρ_A say, at a height $\frac{1}{2}H^*$ above perigee is less than at perigee by a factor $\exp(-H^*/2H)$. Now, if H^* does not differ from the true value H by a factor of more than 1·5—an adequate margin of safety—the value of the quantity $\sqrt{(H^*/H)}\exp(-H^*/2H)$ may be taken as 0·593, with error less than 2·5%. On making this numerical substitution we find

$$\begin{aligned}\rho_A &= \rho_P \exp\left(-\frac{H^*}{2H}\right)\\ &= -\frac{dT}{dt}\frac{0{\cdot}158\ FSC_D}{m}\sqrt{\left(\frac{e}{aH^*}\right)}\left(1-2e-\frac{H^*}{8ae}\right).\end{aligned}$$

The values of air density obtained from this equation, for all the satellites of 1957–9 whose orbits are known and of eccentricity less than 0·2, provide a consistent and reliable picture of the average density at heights between 200 and 700 km, as is shown in fig. 5. At heights above 400 km it is found that the air density depends on the position of the sun, varying gradually between a maximum value when the sun is high, near midday, to a minimum value at a time not far from midnight. The curve of fig. 5 therefore has two branches labelled ‘ midday ’ and ‘ midnight ’ at heights above 400 km.

In the cluster of 27 points between heights of 180 and 300 km, some scatter is to be expected, since the values of m/S have a standard deviation of perhaps 10%. But the scatter is in fact rather greater, the standard deviation being about 20%, chiefly because the density has decreased slightly since 1958. The 24 points at heights between 180 and 250 km can be split into two groups, the first for dates between October 1957 and January 1959 (14 points), and the second for August 1959 to March 1960 (10 points). Curves drawn through each

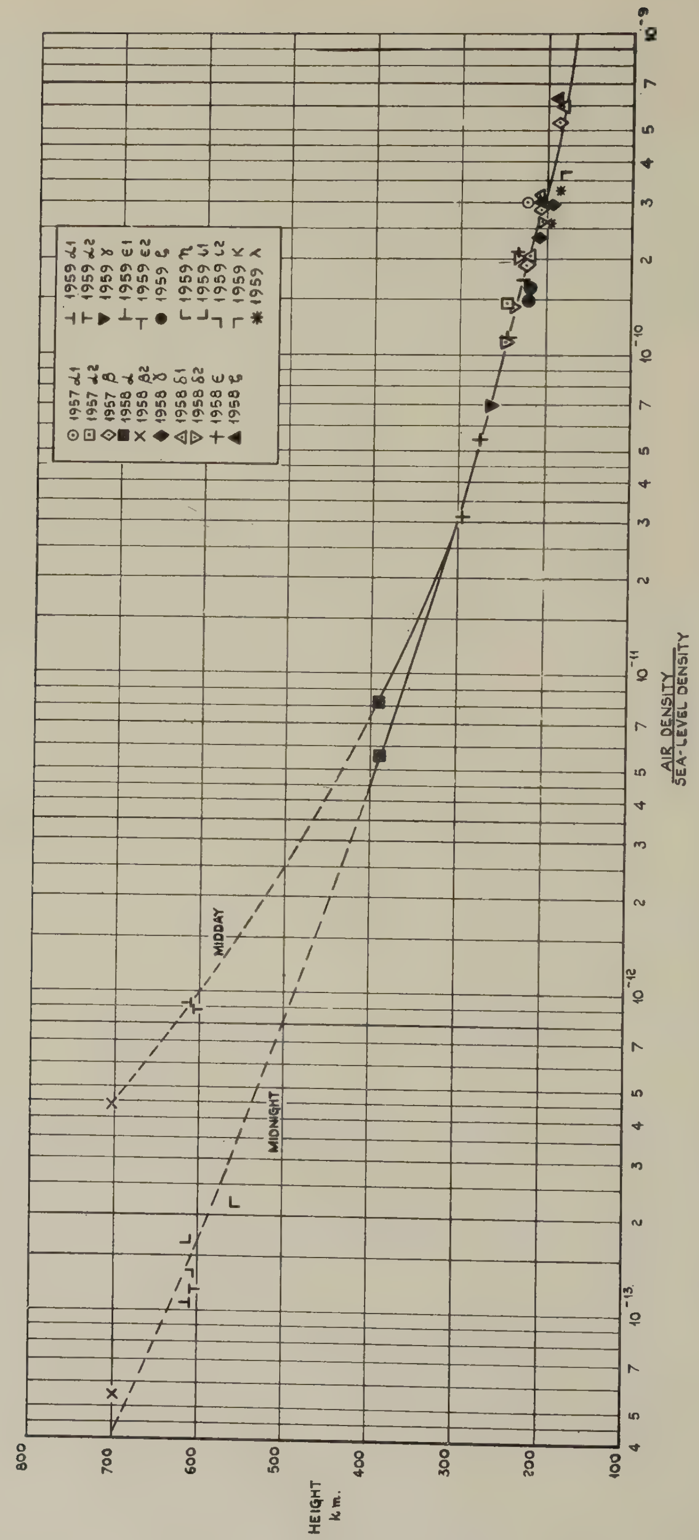


Fig. 5. Variation of air density with height, as revealed by the orbits of 21 satellites of 1957–59. (1957 α = Sputnik 1, 1957β = Sputnik 2, etc.)

of these groups separately give densities which are respectively about 10% higher and 10% lower than the curve of fig. 5, thus indicating that the density has decreased by perhaps 20% between, say, mid 1958 and early 1960. This is interpreted as a variation in density during the course of the 11-year sunspot cycle, with maximum density occurring at the sunspot maximum in 1957–8; a further decrease is expected in the next few years.

The values obtained for air density are liable to rather greater error at heights above about 500 km, because some of the drag on the satellite could be caused by the electrical forces resulting from the passage of a charged body through what is at these heights a quite strongly ionized medium: at heights above 800 km over 10% of the air molecules (or, rather, atoms) may be ionized. It is difficult to make any estimate of the electrical drag because the potential of the satellite is not usually known, but at the moment it appears likely that the error arising from this cause is small, not more than about 10%. The radiation pressure of sunlight can also influence the orbital period of a satellite, but the effect is negligible at heights below 1000 km, except for balloon-type satellites, like Echo.

Before the era of satellites, ideas on the air density at high altitudes were based on a very few, and as it now seems, somewhat inaccurate measurements from research rockets. The atmospheric model generally accepted in 1957 before the advent of satellites gave values of air density which were far too low: the density at a height of 200 km has proved to be three times greater than was previously believed; and at 400 km it is 10–15 times greater. The new and higher values of density also imply that the temperature in the upper atmosphere is several hundred degrees higher than was previously believed. These new findings have invalidated many previous theories of the upper atmosphere and have thrown the subject into an invigorating state of confusion, which it is too early yet to clarify.

5. VARIATIONS IN AIR DENSITY

Any variations in upper-air density are of great interest because they usually reflect some external influence, and so give us a clue to the forces which control the properties of the upper atmosphere.

The method for calculating the average air density can also be used for finding day-to-day changes in density, merely by taking the daily values for the change in orbital period, rather than the changes during a month or more. Another method of detecting these variations is to compare the observed values of $\mathrm{d}T/\mathrm{d}t$ with the theoretical ones, obtained on the assumption that density is constant from day to day. From the equation for T at the end of section 2, the theoretical $\mathrm{d}T/\mathrm{d}t$ is given by

$$\frac{\mathrm{d}T}{\mathrm{d}t} = -\frac{3e_0 T_0}{4t_{\mathrm{L}}\sqrt{(1-t/t_{\mathrm{L}})}}$$

The theoretical curve gives a smooth variation with time, but it was found early in 1958 that the observed values of $\mathrm{d}T/\mathrm{d}t$ for Sputnik 2 were very far from smooth and departed from the theoretical curve by up to 20% or 30%.

At first these variations seemed to be mere irregularities, without any significant pattern. But, as 1958 wore on, it became apparent that they exhibited a 28-day periodicity. Figure 6 is a composite graph made up of results from Sputnik 2 and the rocket of Sputnik 3, which refer to heights of 200–250 km,

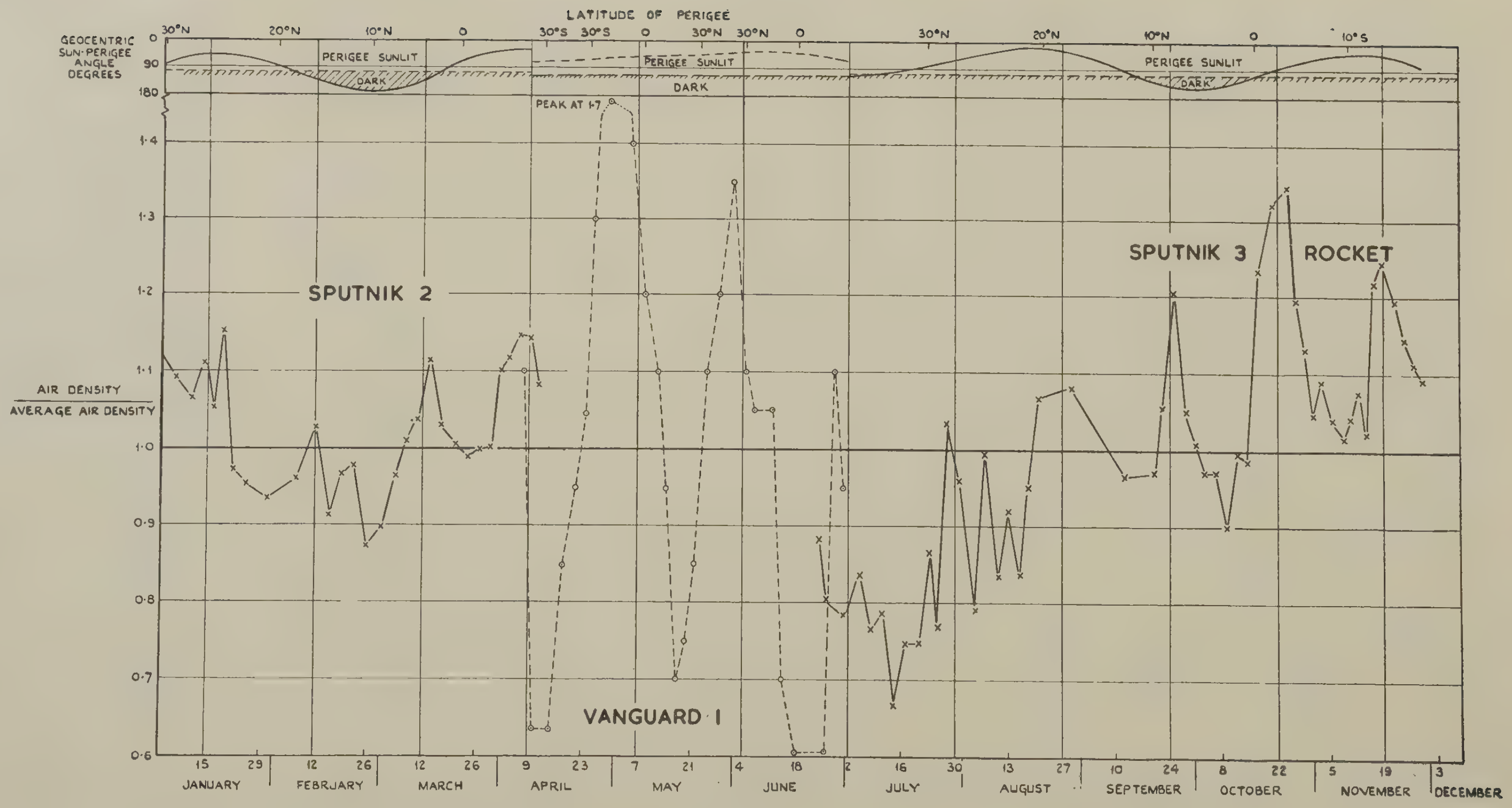


Fig. 6. Variation in air density at a given height during 1958, as revealed by Sputnik 2 and Sputnik 3 rocket (for height of 200–250 km) and Vanguard 1 (650–700 km.)

with the gap between them linked by results from Vanguard 1, at a height of 650–700 km. This graph shows strikingly how the air density during 1958 tended to fluctuate once every 28 days: the grid lines are at 28-day intervals and the maxima of density occur near them. The amplitude of the fluctuations is about ± 20% at a height of 200–250 km and ± 50% at a height of 650–700 km. The results from Vanguard 1, despite their larger amplitude, fit into the pattern well, though they are for a different height and latitude, thus strongly suggesting that the phenomenon is world-wide and extends over a great depth of the atmosphere. (The significance of the scales at the top of fig. 6 will be discussed later.)

It is well known that the sun shoots out streams of charged particles, which swing round with it rather like the spray from a garden water-sprinkler. The sun rotates, relative to the earth, once every 27–28 days, and these streams therefore tend to sweep across the earth at intervals of 27 or 28 days, giving rise to well-known periodicities in the aurora, geomagnetic records and cosmic rays. It thus seemed likely that the periodic variations in the density of the upper atmosphere provided another example of the same phenomenon; and this surmise has been amply confirmed by detailed comparisons between the fluctuations in the drag of the satellites and the solar radiations on various wavelengths. Maximum density occurs when solar radiation and sunspots are at their maximum: so the variation of density in the course of a sunspot cycle, which has already been mentioned, is to be expected.

The mechanism by which the sun exerts its control over the upper atmosphere has not yet been elucidated, though it appears most probable that heating by very short wave radiation is mainly responsible. Whatever the mechanism proves to be, these results have added a new lustre to the theory advocated by Professor Chapman that the earth is, as it were, immersed in the sun's atmosphere, with the earth's upper atmosphere constantly receiving energy from the outer fringes of the solar corona, which, even at the earth's distance, is still at a high temperature, possibly as high as 250 000°K.

At heights of 200–300 km, solar disturbances have proved to be the most pervasive and important influence on air density; but it had at first been expected that there would also be large variations with latitude and season, since direct measurements with research rockets had suggested that such variations existed, even to the extent of a factor of 5 or perhaps 10. The results from satellites have given no support to the idea of any large and systematic variation with latitude or season, however. As fig. 6 indicates, the combined variations from the average values due to solar disturbances, latitude, season and all other causes do not seem to exceed ± 30% at heights of 200–300 km.. and the main variations seem to have no connexion with the latitude scale which is included at the top of fig. 6. At heights above 300 km the variations of density are greater, but there is no reason to suppose that they depend on latitude.

It might also be expected that the air density would be different on the sunlit and dark hemispheres of the earth. At heights between 200 and 300 km, however, this effect has proved to be very small: the diagram at the top of fig. 6 shows that the density is not much affected when the perigee goes from sunlight into darkness, or vice-versa. A slight effect can be detected, however, with the density departing from its average value, as a result of day-to-night variations, by rather less than 10%, and being higher by day. As height increases, the day-to-night variation becomes more important, as fig. 5 shows. At a height of 400 km, the density seems to be about 30% above average at midday and 30% below average near midnight; at a height of 700 km, the maximum daytime value is nearly 10 times greater than the minimum night-time value.

We can best sum up the sources of variation in air density by dividing the relevant height–range into two parts, 180–300 km, and 300–700 km. At heights of 180–300 km, the variations, which are of amplitude ± 20% at the year of sunspot maximum, are closely correlated with solar activity. Variations with latitude and season, and from day to night are much less important, but a significant variation in the course of the sunspot cycle has been detected, density being highest at sunspot maximum (1957–8). At heights above 300 km, the amplitude of the variations increases. The day-to-night effect becomes dominant, with the density by day exceeding the density by night, by a factor which increases from 1·7 at 400 km to perhaps as high as 10 at 700 km. The dependence on solar activity, including the 28-day fluctuation, is also apparent at the greater heights, and causes the density to depart from its average value by a factor of up to about 1·5 at sunspot maximum.

6. AIR TEMPERATURE AT HEIGHTS OF 200–700 KM

The temperature of the upper atmosphere can in principle be found from the rate at which the air density decreases as the height y increases. If the density ρ is proportional to $\exp(-y/H)$, as already assumed, the parameter H can be determined by measuring the slope of the curve in fig. 5: H is the height in which density falls off by a factor of 2·718, and its average value at heights between 200 and 400 km is about 50 km. The air temperature in degrees Kelvin can then be evaluated with adequate accuracy from the expression

$$\frac{1{\cdot}18\, MHg}{(1-H')\,g_0} \quad {}^{\circ}\mathrm{K},$$

where H is in kilometres, H' is the rate of increase of H with y, M is the mean molecular weight of the air, and g and g_0 are the local and sea-level values of the acceleration due to gravity.

Unfortunately neither the value of H nor that of M is known at all exactly: measuring the slope of the curve in fig. 5 is obviously a process fraught with error, and the exact composition of the air at great heights is a subject of controversy. From fig. 5 and other evidence, H can be taken as varying linearly between about 40 km at a height of 200 km and 70 km at 400 km height, with a standard error which might be as high as 20%. At a height of 150 km the air is largely composed of molecular nitrogen (M.W. 28) and atomic oxygen (16) in much the same proportions as at sea level, so that M is about 25. At a height of 400 km the picture is much less certain, but it is probable that atomic oxygen predominates, with a much smaller amount of nitrogen, in either molecular or more probably atomic form, so that M is not likely to be much greater than 16. If these values of H and M are correct, the temperature increases from about 1150°K at a height of 200 km to about 1400°K at a height of 400 km. At heights above 400 km the temperature is greater by day than by night: at 600 km the values are about 2200°K by day and 1200°K by night.

7. WINDS IN THE UPPER ATMOSPHERE

If the atmosphere did not rotate, the aerodynamic force on a satellite would act in the orbital plane; in fact, however, the atmosphere rotates with the earth, and subjects the satellite to a ' side-wind ', which in turn produces a small force perpendicular to the orbital plane. The net effect of this force is slightly to change the inclination of the orbit to the equator. The magnitude of the effect depends on the mean rotational speed of the atmosphere near perigee. So, by comparing the observed and theoretical values for the change in inclination we should be able to deduce the mean wind speeds.

In practice, however, the accuracy of the method has been poor, because the changes in inclination are so small. For Sputnik 2 the theoretical change in inclination during the lifetime, for zero wind, was a decrease of 0·12°: the observed decrease was from 65·33° to 65·19°. For the rocket of Sputnik 3 the theoretical decrease was 0·13°: the observed decrease was from 65·18° to 65·00°. The observational values are therefore greater than those predicted for zero wind, thus suggesting that a strong west-to-east wind exists at heights of 200–250 km. This suggestion must at present be treated as rather conjectural, since the errors in the observed values may be as large as 0·02°. In principle, however, the method is a useful one, and when more accurate observations become available as a matter of routine, each new satellite should yield information on the wind speeds at heights near that of its perigee.

REFERENCES FOR FURTHER READING

Satellites and Scientific Research by D. G. King-Hele (Routledge & Kegan Paul, 1960) describes the scientific results obtained from satellites more fully. An excellent background picture of the earth and its atmosphere is provided in *Once Round the Sun* by R. Fraser (Hodder & Stoughton, 2nd ed., 1959) and *The Upper Atmosphere* by H. S. W. Massey and R. L. F. Boyd (Hutchinson, 1958). For more technical review articles on the topics covered in this article, see ' Developments in Dynamical Geodesy ' by A. H. Cook, *Geophysical Journal of the Royal Astronomical Society*, Vol. 2, pp. 222–240 (1959), and ' Properties of the atmosphere revealed by satellite orbits ' by D. G. King-Hele, *Progress in the Astronautical Sciences*, Vol. 1 (1961).

The Author:

Is a Senior Principal Scientific Officer at the Royal Aircraft Establishment, Farnborough, where in the last few years he has worked mainly on earth-satellite orbital theory and interpreting observations to derive geophysical information. He is author of many papers on these topics and of two books published in 1960, *Satellites and Scientific Research* and *Shelley: his Thought and Work.*

Interference Spectroscopy

by H. J. J. BRADDICK
University of Manchester

Spectroscopy has for many years been one of our chief sources of information about the constitution of matter, and a very important tool in applied science. Prisms and diffraction gratings have usually been used for sorting the radiations, and the photographic plate for detecting and recording them, at least in the visible and ultra-violet parts of the spectrum. The prism and grating have been supplemented by interferometers when high resolution has been required to examine the detailed structure of spectrum lines, or when wavelength measurement has been pushed to the highest accuracy.

Photoelectric cells have sometimes been used to detect radiation in spectroscopy, and in the post-war years photoelectric methods have gained ground rapidly, largely because electron-multiplier photocells (' photomultipliers ') offer great advantages of convenience and sensitivity. The fundamental advantages of the photoelectric methods of detection lie in the high quantum efficiency of the photocathodes and in their linear response to radiation intensity; against this must be set the fact that the photographic plate can record simultaneously data from a number of elements of a spectrum. We must regard the information contained in a spectrum as consisting of the intensities at a number of points (sometimes called ' picture points '). It is sometimes the object of the spectroscopist to obtain the positions of a large number of spectral lines; from our present point of view crude intensity values are then required at a large number of picture points, but in an increasing number of applications, technical and purely scientific, the information sought is given by the intensities at only a few points. For example, the analysis of an alloy can be obtained by measuring the relative intensities of a limited number of spectral lines; the effective electron temperature in a cloud of interstellar gas may be obtained from the ratio of the intensities of two or three spectral lines properly chosen. In such cases photoelectric photometry is advantageous.

It is also true that in many spectroscopic problems the source of light has a low intrinsic brightness, so that after analysis of the light by a spectroscope, only a small flux of radiant energy is available for each picture point. Now in such cases the accuracy of the intensity measurements is limited by ' noise ', in the sense of statistical fluctuations arising either from the random arrival of photons at the detector, or from random disturbances inherent in the detector itself. These fluctuations can always, in principle, be ironed out by averaging the measurements over a period of time, but the time required may be quite long and indeed for some reason it may not be available. The efficiency with which the spectrograph collects and transmits light, as well as the quantum efficiency and the internal noise of the detector, may therefore come to have primary importance in research.

Photocell detectors, together with thermocouples and other infra-red detectors, measure directly the whole flux of radiation falling upon them; the analysis given below shows how this quantity depends on the type of spectroscope used. It must be remembered that the photographic plate responds primarily to *flux per unit area* and the human eye responds primarily to *brightness* which depends on the flux divided by the area of the image and the solid angle of the beam entering the eye. The analysis is therefore not immediately applicable to these detectors.

In accordance with a very fundamental theorem of optics, the maximum flux of radiation which can be passed through an instrument such as a spectroscope, which has a well defined entrance aperture, depends directly on the brightness of the source, the area of the entrance aperture, and the solid angle of the beam passing through every point of the aperture. Now, with a given dispersing system, the spectral resolution of a spectroscopic instrument is connected with the dimensions of the entrance aperture; if we have a given spectroscope and a source of given monochromatic brightness we can only supply an increased flux to the detector at the expense of resolution, as by widening the slit of a prism or grating spectroscope. (It should be pointed out that widening the slit confers no advantage if the detector is a photographic plate and we are examining a weak line spectrum. The increased flux is spread over a larger area of the plate and the conditions for photographic response are no more favourable.)

Jacquinot (1954) considered systematically this aspect of spectroscope design and showed that in certain interferometers the relation between spectral resolution and the transmission of luminous flux could be much more favourable than in conventional spectroscopes. This result has been applied to a number of spectroscopic problems (Bellevue Conference 1957) and is likely to find more applications in future spectroscopy both scientific and technical.

The simplest interference spectroscope is a Fabry–Pérot étalon consisting of two accurately parallel transparent flats coated with efficient (i.e. non-absorbing) partially reflecting layers and used with light incident normally. Such an étalon acts as an interference filter transmitting light only in a number of very narrow spectral regions centred on the wavelengths given by $n\lambda_n = 2\mu D$ where D is the separation of the surfaces, μ is the refractive index of the medium between them and n is an integer. It is convenient to call $2\mu D$ the optical thickness of the étalon.

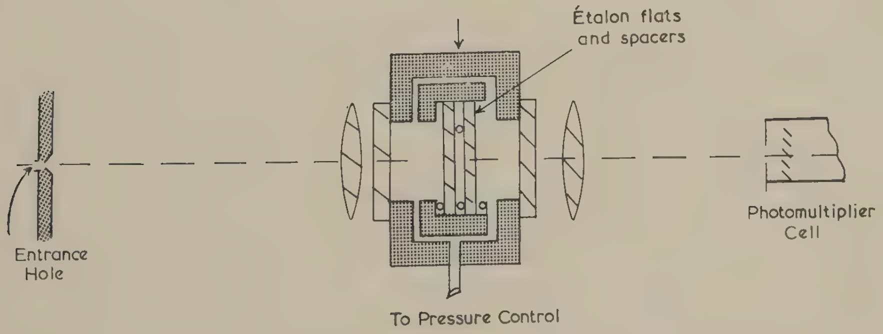


Fig. 1. Diagram of a Fabry–Pérot interferometer arranged for scanning by variation of gas pressure.

By using an auxiliary monochromator, or otherwise, we can arrange to work with only one of the maxima, and by varying the optical thickness we can make the transmission band scan a certain spectral region. An intensity record of the region can be obtained from a photoelectric or other flux detector which receives the radiation after it has passed through the étalon. Several methods have been suggested for varying the optical thickness of an étalon; the most effective ways as yet seem to be accurate mechanical displacement of the interferometer plates or changing the pressure of a refracting gas between the plates (fig. 1). Using air between the plates, one atmosphere pressure corresponds to a change in refractive index of about 3×10^{-4} and allows the exploration of a fractional spectral range of about 1/3000, but there are gases such as trifluorochloromethane (9×10^{-4} per atmosphere), and propane (14×10^{-4} per atmosphere) which are considerably more refracting than air and which give correspondingly larger scans for a given pressure change. The pressure change which may be used is limited only by the design of the apparatus; it is of course necessary to avoid gases which absorb light appreciably in the spectral region under investigation.

Following Jacquinot, we shall demonstrate the flux transmitting properties of different spectroscopes by considering first an ideal blazed grating and then a Fabry–Pérot interferometer. In each case the dispersing element, grating or interferometer plates, has area S. Light is admitted to the grating spectroscope (fig. 2*a*) by a rectangular slit of height h and width W, and in the most efficient arrangement this slit is covered by the monochromatic brightness B of the source. The light is collimated by a lens or mirror of area S and focal length f before falling on the grating. The flux incident on the grating and finally on the detector is given by $L = BSt\,hW/f^2$ where t is the transparency of the system and a small-angle approximation is used for the solid angle.

The resolution is supposed to be limited by the finite width of the slit rather than by diffraction, and this is usually true in the practical spectroscopy of weak light. If $\delta\lambda$ is the smallest resolvable wavelength interval we have $\delta\lambda = d\lambda/d\theta.\ W/f$, since W/f is the angular width of the slit and $d\lambda/d\theta$ is the reciprocal of the angular dispersion of the grating. If the incident and emergent rays make angles ϕ, θ, with the grating normal we have

$$\frac{1}{\lambda}\frac{d\lambda}{d\theta} = \frac{\cos\theta}{\sin\theta + \sin\phi}$$

which is of the order of unity for practical spectroscopes; $R = \lambda/\delta\lambda$, which we take as a measure of resolution, is therefore approximately f/W. The flux L is now seen to be $(BSth/f)(1/R)$, which displays the inverse relation between flux and resolution under the stated conditions. The ratio *flux/resolution*, given unit source brightness and unit area of the dispersing element, may be taken as a figure of merit characteristic of a type of spectrograph. For an ideal grating it is given as h/f, which in practice is 1/15 or less for most grating spectroscopes.

For the Fabry–Pérot interferometer used as a monochromator with a beam parallel to the axis (fig. 2*b*) the entrance aperture would be a circular hole of radius r placed at the focus of a collimating lens. The spectral resolution is now ideally limited by the finite range of angles of incidence admitted by the collimator, for when light of wavelength λ is passed normally by the étalon of thickness D, and wavelength $\lambda - \delta\lambda$ is transmitted in the same order at angle θ, we have $n\lambda = 2\mu D$, and $n(\lambda - \delta\lambda) = 2\mu D \cos\theta$ so that

$$R = \frac{\delta\lambda}{\lambda} = (1 - \cos\theta) = \frac{\theta^2}{2} = \frac{1}{2}\frac{r^2}{f^2}\,.$$

(The effective resolution is rather greater than this since the solid angle corresponding to $\theta = 0$ is zero and the effective range of wavelength transmitted is smaller than that calculated.) The flux through the system is $\pi\gamma^2 BSt/f^2$ and the figure of merit, calculated as above, is 2π. This is much larger than the figure of merit obtained for the grating, and might in some applications be further increased by using an annular diaphragm as well as the central hole, thus admitting more than one order of the interference pattern.

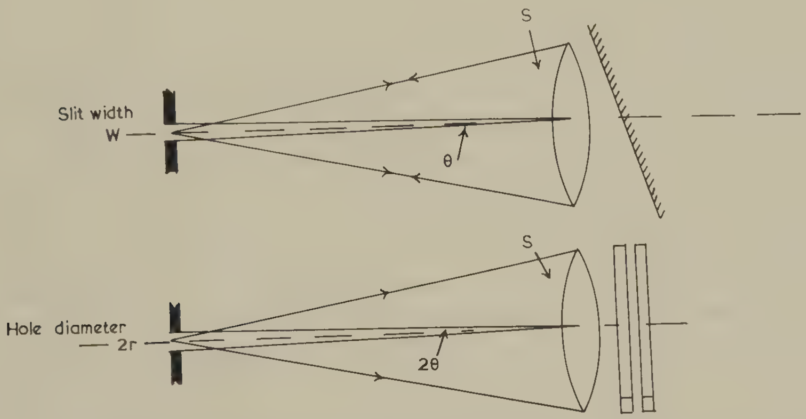


Fig. 2. Comparison of the luminous efficiency of spectrometers. (*a*) Grating spectrometer, drawn as an autocollimating system; (*b*) Entrance side of Fabry–Pérot system.

This superiority is clearly due to the fact that the phase difference between interfering beams within the étalon varies only slowly when the point of incidence of a ray moves in any direction across the plane of the entrance diaphragm, while the phase difference produced by a small change in wavelength is large; this is equivalent to saying that the linear dispersion, measured in any direction across the diaphragm, is large, and the dispersion is in fact infinite on the axis of a Fabry–Pérot.

In prism and grating spectroscopes, the effect of moving the point of incidence along the slit is small, but the wavelength transmitted by the instrument changes rapidly if the point of incidence moves across the slit, and high wavelength resolution requires a narrow slit. As is to be expected from the smaller angular dispersion usually achieved with a prism system, the figure of merit of a prism spectroscope is usually much smaller than that of a grating.

The Fabry–Pérot interferometer arrangement has rotational symmetry and infinite dispersion when used on the axis: some other interferometers such as the Lummer plate and the Fabry–Pérot used with oblique incidence do not have this symmetry and the figure of merit is much lower. The Michelson two-beam interferometer can be used axially and has then similar properties to the axial Fabry–Pérot with respect to intensity, though the way in which it obtains information from a spectral distribution is distinctly different.

The geometrical figure of merit does not of course tell the whole story; the flux transmitted from a source of given monochromatic brightness depends also on the transparency of the spectroscope and the size of the dispersing element. The transmission of a good modern blazed grating at the maximum of the blaze is 50 to 70 per cent. The peak transmission of a Fabry–Pérot interferometer with multiple dielectric films as reflecting coatings is at least as good as that of a grating. The maximum size of the dispersing element is largely an economic matter but again there is no overwhelming difference between the instruments. The interferometer therefore retains in practice an advantage over the grating which is comparable with that calculated from the geometrical figures of merit. It is of course necessary that the size and shape of the source shall allow the entrance aperture to be filled in each case.

The significance of improved luminosity in a spectroscopic instrument is apparent when the source brightness is so low that fluctuation phenomena (noise) become important. There is a fundamental limitation to the accuracy of intensity measurement which arises from fluctuations in the rate of detection of photons. If n is the number of photons which fall on the detector each second, ϵ is the quantum efficiency of the detector defined by the average number of *independent detectable events* produced by each photon, and if observations are integrated over time τ, the ratio of the measured quantity (signal) to the r.m.s. fluctuation is $n\epsilon\tau/\sqrt{(n\epsilon\tau)} = \sqrt{(n\epsilon\tau)}$. In the best photoelectric detectors amplification and recording cause very little deterioration in this signal-noise ratio.

If, therefore, we use a spectroscope and a given flux detector to measure to a given accuracy the spectrum of a given source, the time required for each picture point is inversely proportional to the flux incident on the detector. As we have seen, once the resolution required has been specified, this flux depends on the figure of merit of the spectroscope and any increase in this figure can lead to a reduction of observation time.

With infra-red detectors at present available the photon fluctuations are swamped by detector noise independent of the incident light. In the ideal case the signal/noise ratio is $n\epsilon\tau/\sqrt{(\eta\tau)}$ where η measures the rate of noise events characteristic of the detector, and the time required for a given accuracy decreases with the *inverse square* of the number of photons falling on the detector, and hence with the inverse square of the figure of merit. This improvement may not be achieved if it is necessary to use a larger detector with increased noise to accommodate the larger exit pupil.

A reduction of observation time can be achieved if more than one picture point can be examined at the same time. This can be done with a photoelectric grating spectroscope if a number of independent detectors are provided: it is not normally possible with a Fabry–Pérot system. It will be shown below that there is an interferometric technique which allows the simultaneous study of a number of spectral elements, but that it only gives an advantageous signal/noise ratio if detector noise and not photon fluctuations set the limit to accuracy. In photographic spectroscopy many elements are of course recorded simultaneously but the response of the plate is not linear and the information about each spectral element is spread over a region on the plate and has to be recovered by using a densitometer. It is therefore difficult to compare quantitatively the efficiency of photographic and photoelectric methods. The quantum efficiency of the photographic plate is known to be low compared with that of a photocathode and it has been found empirically that the information gathered in a given time by a photographic plate may be compared with that obtained by applying a photocell successively to about 100 spectral elements. This allows a rough comparison to be made between, say, a photographic grating spectroscope and a photoelectric interferometric instrument. It is clear that the latter instrument is likely to be better unless a large number of picture points is required. If this advantage is to be secured, time must not be wasted in scanning unnecessary spectral elements, and, as remarked above, the source and optical system feeding light into the interferometer must fill the entrance aperture with light.

It is notorious that interference spectroscopes have a small free spectral range; there is only a small wavelength separation between the successive orders corresponding to successive values of n in the equation $n\lambda_n = 2\mu D$ in the case of a high-order étalon used as a chromatic filter. It is therefore necessary to use a spectroscope of lower resolution as a preselector, and to arrange matters so that this instrument does not reduce seriously the luminous efficiency of the assembly. Unfortunately there is a limiting relation between the free spectral range of an étalon and its resolution which decides the minimum performance of the auxiliary monochromater. The ratio between the fringe width of an ideal monochromatic line, measured at half intensity, and the separation of successive orders is conveniently called the sharpness of the fringes; it depends on the reflecting power of the surfaces and on their optical flatness, and the best values attained in present practice lie between 20 and 100. This means that the resolution of the pre-selector cannot be less than, say, one fiftieth of that of the étalon, and if a grating is used as a pre-selector this reduction in resolution does not offset the lower figure of merit. The grating spectroscope is then likely to lower the luminous efficiency of the combined system.

High efficiency can be obtained at the expense of complication by using an étalon of intermediate resolution between the grating spectroscope and the final étalon.

In order to cover an appreciable spectral range, it is necessary to scan simultaneously on the final étalon, the pre-selector and the intermediate étalon if there is one. As has been implied on p. 270 there is difficulty in scanning an étalon over an extended range. It is however possible, having scanned a certain region by simultaneous variation of the étalon and pre-monochromators, to reset the étalon, keeping the pre-monochromator temporarily stationary so that an adjacent order of interference takes over the frequency which is being transmitted by the monochromator. The étalon scan can now begin again, the monochromator keeping in step. This process, called saw-tooth scanning, can be repeated indefinitely, but all this complication is required only if we wish to analyse an extended and complicated spectrum. In some of the problems which are most rewarding for interference spectroscopy the investigation is confined to isolated lines of an emission spectrum and the pre-selecting monochromator can have very low resolution.

A problem which shows the method at its most advantageous is the exploration of the widths of lines in the light of the night sky, leading to estimates of the effective temperature of upper atmospheric regions. In this work the source is so extended that there is no difficulty in filling the interferometer pupil, and a simple filter is all that is required as a pre-selector. Another problem to which the scanning interferometer has been applied is the isotopic analysis of uranium by measuring the intensities in the hyperfine structure of a line, different components arising from different isotopes.

In the applications already mentioned, the resolution required lies in the region conventionally appropriate to interferometers, but there are practical cases in which the resolution could easily be obtained in other kinds of spectroscopes and the interferometer is used solely on account of its luminous efficiency. In any large or medium-sized astronomical telescope, the image of a star is not the diffraction disk of simple theory, but a ‘ seeing disk ’ due to atmospheric disturbances. The light of the star is spread over several seconds of angle and the entrance aperture of a spectroscope has to be fairly large if it is to collect light efficiently in the focal plane of the telescope. Furthermore, in a spectroscope which works by time-scanning, serious irregularities are introduced into the spectrum by fluctuations in the size and position of the stellar image unless nearly all the light goes through the spectroscope aperture all the time. It is also true that some astronomically interesting objects are nebulae of appreciable angular size. There is therefore a good case for using interference spectroscopes on astronomical telescopes. Work of this kind has only just begun, but it has been possible to obtain, using a telescope of moderate size, measurements on the lines of ‘ planetary ’ nebulae which would have been difficult to get in any other way.

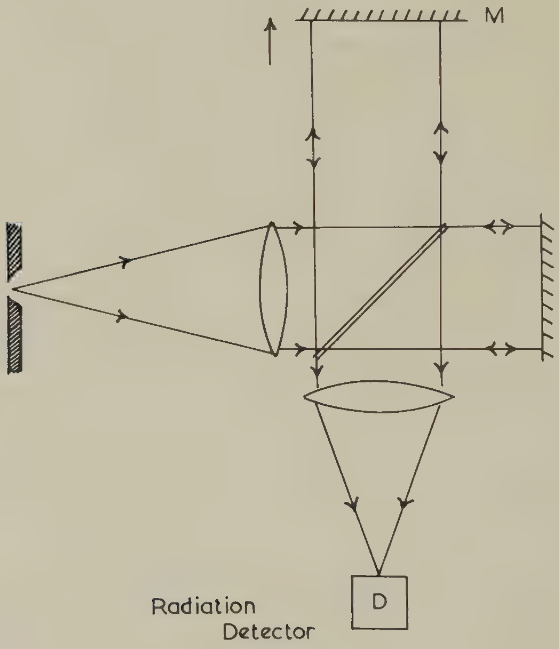


Fig. 3. Arrangement of the Michelson interferometer for Fourier spectroscopy. The mirror *M* is moved steadily in the direction shown.

In infra-red spectroscopy, photographic methods are not available, and flux detectors with time-sequential scanning must perforce be used. The accuracy of measurements in this region is usually limited by the weakness of the radiation, and interference spectroscopes may therefore have important applications. The development of efficient infra-red reflecting coatings for Fabry–Pérot plates is still in a rather early stage of development. It appears that semi-conducting films may be alternated with dielectric films to form suitable coatings.

In the infra-red it is also possible and advantageous to employ an interference method using only two interfering beams. Such interference may be obtained in Michelson's well-known beam-splitting interferometer, and the new method is an extension and refinement of Michelson's own method of studying the 'visibility' of the fringes. If the Michelson interferometer is illuminated with collimated light and the mirrors are kept effectively parallel, the interference fringes can be observed in time sequence by measuring the output intensity while one of the mirrors is moved so as to vary the optical path (fig. 3). The symmetry and luminous efficiency of this arrangement are similar to those of the Fabry–Pérot interferometer. If the radiation observed were strictly monochromatic the variation of intensity with path difference would be sinusoidal. In an actual source the energy for each wave-number interval $d\nu$ is given by a function $E_\nu\, d\nu$, and the output intensity due to the frequencies in this range is $E_\nu\, d\nu \cos^2(\delta/2) = E_\nu\, d\nu\,(1+2\cos 2\pi\nu x)$ where δ is the phase difference between the beams and x is the corresponding path difference. As δ is varied the intensity variation of the whole emergent flux can be recorded by a suitable instrument. It is given by the Fourier cosine transform

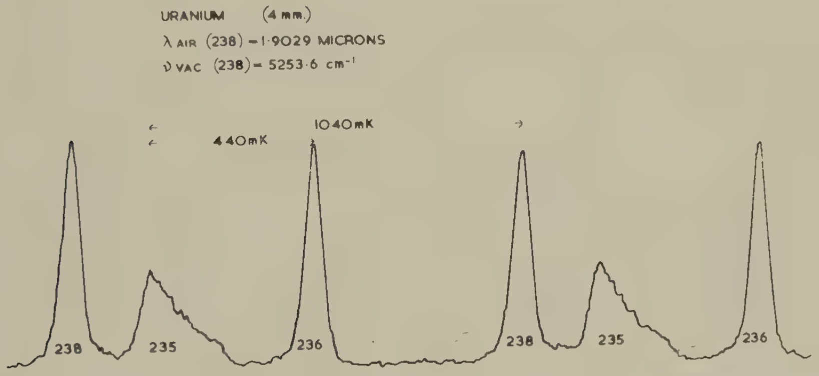


Fig. 4. Pressure scanned interferometer record of the isotopic structure of an infra-red line of uranium. (R. Beer, Manchester and L. Bovey, Harwell.)

$$F(x) = \int_0^\infty E_\nu \cos 2\pi\nu x\, d\nu$$

and the original spectral distribution can be obtained by carrying out the inverse Fourier transformation

$$\int_0^\infty F(x) \cos 2\pi\nu x\, dx.$$

This operation can be done by an analogue or a digital computer. The original spectrum can be fully recovered, subject to a limitation of resolution which is mentioned below. Michelson himself did not measure the output from his interferometer, but estimated by eye the contrast or ' visibility ' of the fringes. In this estimation, the information corresponding to the relative phase of the Fourier components is lost, and ambiguity is introduced into the reconstruction: It is, however, well known that Michelson's work gave the earliest indications of the structures of certain spectral lines.

For an accurate reconstruction of a spectral distribution it would be necessary to have the function $F(x)$ for all path differences to infinity; in practice only a finite range of path difference can be used and this limits the resolution obtainable. For example the interferometer output derived from a monochromatic source is a rectangular block of sine waves instead of a sine function extending with constant amplitude to infinity. The corresponding inverse Fourier transform shows a spurious structure: it is in fact a distribution of the form $\sin^2 \phi/\phi^2$ instead of a sharp line. It is found that the resolution $\lambda/\delta\lambda$ is of the same order as the number of wavelengths in the change of path difference which constitutes the scan.

In this method of spectroscopy the whole of the spectral input contributes to every point of the output, so that the efficiency with which information is collected may be expected to be high. The photon fluctuation noise associated with the whole spectrum is, however, observed during the whole time of the scan. This usually means that the signal-noise ratio is poor if the photon fluctuations are the dominant source of disturbance, and the usefulness of the method in the visible spectrum is confined to special cases. In the further infra-red part of the spectrum the dominant noise arises from the detector rather than from the incoming photons, and the two-beam method is useful. In this part of the spectrum the resolution which can be applied, using conventional spectroscopes, is often limited in practice by the low intensity of radiation available. Effective reflective coatings for Fabry–Pérot interferometers are not yet obtainable and the Fourier–Michelson method is the best available: its efficiency has been demonstrated mainly in investigations of atmospheric absorption. There is another field of application, very close to Michelson's own work, in which the two-beam interferometer is valuable. This is the detailed investigation of spectral lines whose approximate structure is known. This is specially important to metrologists who wish to examine different spectral lines and different ways of producing them, so that they may be used as standards of length.

The main source of information on the subject is still the report of the 1957 Bellevue Conference (*Journal de Physique*, **19,** 187. March 1958), which contains most of the work described above and some other important devices of rather less general application. Several workers have now applied the methods of interference spectroscopy, mostly in the fields of investigation which have been mentioned. Since this review was written the subject has been treated by Jacquinot in *Reports on Progress in Physics*, 1960.

The Author:

Dr. H. J. J. Braddick is Reader in Physics in the University of Manchester. He has written *The Physics of Experimental Method* as well as scientific papers. He is at present interested in developing and applying the methods of spectroscopy described in this article.

The Microtron

by R. E. JENNINGS
University College, London

1. Introduction

The microtron is an accelerator for electrons which was suggested by Veksler in 1945. The acceleration is achieved by using a resonant cavity which is situated in a constant magnetic field, the cavity being fed from a pulsed high powered r.f. source at a wavelength of a few centimetres. Such sources were developed during the war and the first microtron, operating at $4\frac{1}{2}$ MeV, was constructed in Ottawa (Redhead *et al.* 1950). A similar machine, with facilities for extracting the beam, was constructed in this country shortly afterwards (Henderson *et al.* 1953 (2)).

In many ways the microtron resembles the cyclotron which is used to accelerate comparatively heavy particles, such as the proton which has a rest mass of nearly 2000 times the rest mass of the electron. Both machines use a steady magnetic field to bend the trajectories of the particles into circular orbits so that an alternating electric field can be used for repeated acceleration of the particles as they cross the same gap. However, whereas the fixed frequency cyclotron runs into difficulties as soon as the relativistic increase in mass becomes at all appreciable, the microtron relies for its operation on a large increase in mass at each crossing of the gap. Before describing in detail the mode of operation of the microtron (or electron cyclotron as it is sometimes called) it will be useful to consider how a cyclotron works.

In a cyclotron (fig. 1) the particles are accelerated each time they cross the gap between two flat semi-circular boxes, called dees on account of their shape. The dees are situated in a magnetic field and the acceleration is caused by applying an alternating voltage between them. After each crossing of the gap, the particle passes into one of the hollow dees where it is screened from the

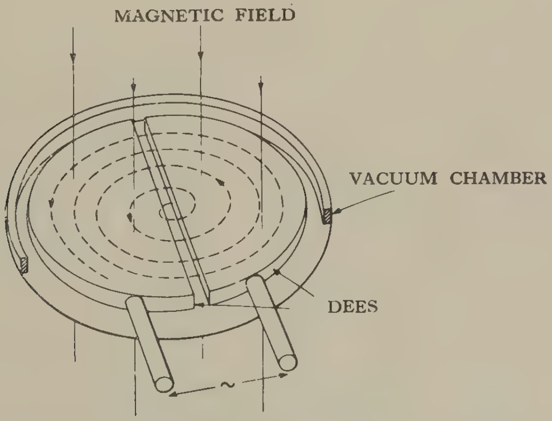


Fig. 1. Basic arrangement of the cyclotron. The orbits are not drawn to scale; in practice the orbits, whose radii increase as the square roots of the natural numbers, are very much closer together.

electric field at the gap and while it follows its semicircular orbit inside the dee the electric field changes sign so that the particle is again accelerated at its next crossing of the gap. For this acceleration to continue, the period of the alternating voltage must be matched to the revolution time of the particle in its orbit. An expression for the revolution time can be derived as follows:

If a particle of charge e (esu) moves normal to a magnetic flux of density gauss with a velocity v it experiences a force of $\frac{Bev}{c}$ in a direction normal to both that of the magnetic field and the direction of motion of the particle. It is this force which makes the particle follow a circular orbit in a cyclotron, and it is equal to $\frac{mv^2}{r}$ where m is the mass of the particle and r the radius of the orbit. From $\frac{Bev}{c}=\frac{mv^2}{r}$, the revolution time $T\left(=\frac{2\pi r}{v}\right)$ is equal to $\frac{2\pi mc}{Be}$. It is interesting to note that the revolution time T is independent of the size of the orbit, the increase in radius as the particle is accelerated being exactly counterbalanced by its increase in velocity, this fact explaining why it is possible for a cyclotron to operate at a fixed frequency. However, this is not the whole story and this type of acceleration cannot continue indefinitely. As the velocity and the energy of the particle increase its mass can no longer be considered constant. When the relativistic mass of the particle, given by $m=m_0/\sqrt{1-v^2/c^2}$ becomes significantly greater than its rest mass m_0 the revolution time T becomes appreciably longer. This increase need only be very small to upset the operation of the machine, as each particle makes a large number of revolutions.

In practice this difficulty is overcome, in the synchro-cyclotron, by slowly decreasing the frequency of the voltage across the gap during the acceleration of each bunch of particles. In principle it could be overcome by using such a large voltage between the dees that the increase in mass of the particles at each transit was sufficient to make them take extra time corresponding to a whole period of the accelerating voltage for each successive half orbit so that they would again be in phase. In the case of the cyclotron, accelerating protons, the voltage required is far too high, being of the order of thousands of millions of volts, but this is exactly the idea behind the microtron. Here the particles being accelerated are electrons, of rest mass about 1/2000th of that of a proton, and the corresponding accelerating field of the order of $\frac{3}{4}$ millon volts/cm can be fairly readily achieved nowadays with high $-Q$ cavities fed from pulsed r.f. sources at a wavelength of a few centimetres. The orbits of a microtron are shown in fig. 2. They have a common tangent along the axis of the resonant cavity. The operating frequency of a microtron is considerably higher than that of a cyclotron due to the smaller mass of the electron.

2. Resonant Conditions

In discussing the resonant conditions, it is convenient to consider the acceleration as being caused by a voltage between the inner lips of the cavity. If the effective voltage is V, then the energy gained per transit by an electron is eV (often measured in electron-volts). There are two conditions for resonant acceleration.

1. Consider electrons starting with negligible kinetic energy on one side of the cavity; these electrons must take a whole number of periods of the r.f. supply to complete the first orbit.

   The time for an electron to complete the first orbit is $\frac{2\pi}{Be}mc = \frac{2\pi}{Bec}mc^2$, where m is the mass of the electron. Here mc^2 is the total energy of the electron and we thus see that the time for an electron to complete an orbit is proportional to its total energy. For the first transit, $mc^2 = m_0c^2 + eV$ where m_0c^2 is the rest energy of the electron which is equal to 0·511 MEV. (1 MEV $= 10^6 eV$).

   The first resonant condition is therefore that $\frac{2\pi}{Bec}(m_0c^2 + eV) = a\tau$ where a is an integer and τ is the period of the r.f. supply.

2. The second resonant condition is that an electron must take an integral number of periods more for each successive orbit, i.e.

$$\frac{2\pi}{Bec}.eV = b\tau \text{ where } b \text{ is an integer.}$$

These two conditions lead to values for B and V.

$$B = \frac{2\pi\, m_0c}{e\tau(a-b)} \text{ and } V = \frac{m_0c^2}{e}\left(\frac{b}{a-b}\right)$$

What are the optimum values for a and b? The magnetic field should be as large as practically possible, so $(a-b)$ is made equal to 1. On the other hand the corresponding voltage should be as small as possible, i.e. $b=1$ and consequently $a=2$.

This mode is the normal or fundamental mode. In this mode, the energy gain per transit is $eV = m_0c^2$, i.e. the rest energy of the electron corresponding to an effective voltage between the lips of the cavity of a little over half a million volts (i.e. after allowance for phase of transit, etc.). In this mode the electron leaves the cavity on its first orbit with a total energy of two rest masses and takes two periods to complete the orbit. It then crosses the cavity again in phase, receives additional energy corresponding to one rest mass, takes three periods to complete the next orbit and so on. Initially the electron's velocity increases rapidly and is soon only slightly less than that of light. Thus, as the acceleration proceeds, the velocity of the electron remains effectively constant but its mass continues to increase. After the first few orbits, each orbit increases in length by almost the free space wavelength λ of the r.f. supply and so the orbits are separated by $\frac{\lambda}{\pi}$ at points opposite the resonator, as shown in fig. 2.

For the normal mode the value of B is 1070 gauss for $\lambda = 10$ cm. This field is unfortunately comparatively low and consequently the radius of a microtron magnet is an order of magnitude larger than the corresponding magnet for a betatron (say). However, as the field is constant, the magnet is much cheaper to construct (it does not have to be laminated) and easier to operate than an a.c. magnet. This is a big advantage.

The field B is proportional to $\frac{1}{\lambda}$ but unfortunately it is difficult to decrease the wavelength much below 10 cm and still operate in the fundamental mode. H. F. Kaiser (1956) has done a considerable amount of development work on microtrons operating at wavelengths below 10 cm in higher order modes.

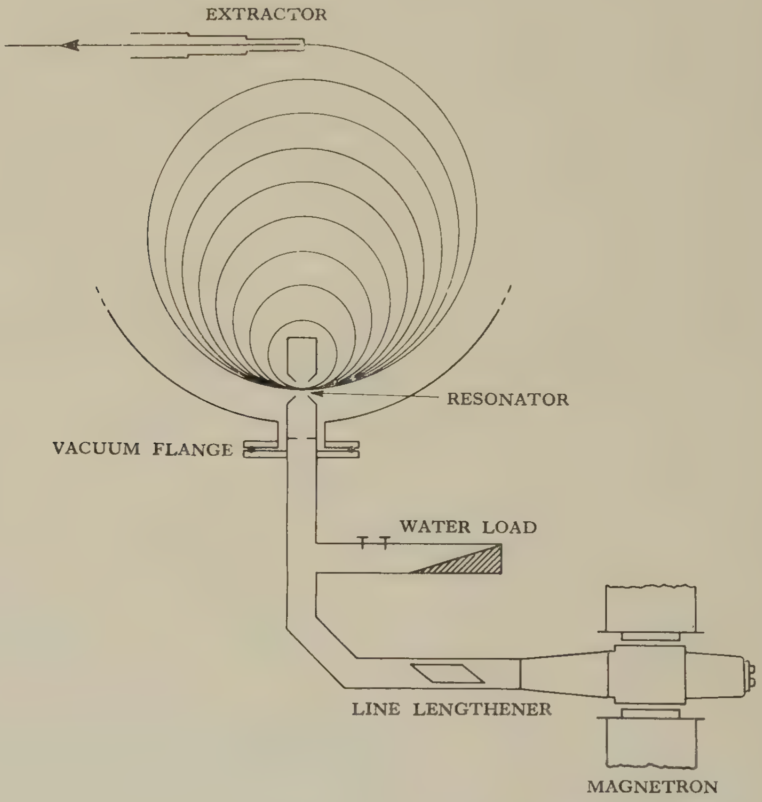


Fig. 2. Orbits and r.f. supply of a microtron.

3. Phase Stability

The behaviour described in the previous section is for an ' ideal ' electron which obtains an energy increase of exactly the correct amount at each transit. If a good beam is to be obtained from the machine it is important that electrons which deviate slightly from the ideal conditions should still be accelerated without loss. In the case of the microtron the calculation to find the extent of permitted deviations has to be made step by step as the energy increase at each crossing of the cavity is so large that the corresponding errors in energy and phase are too large to be represented by differential equations.

The region of phase stability for a microtron is after the peak of the r.f. field in the cavity. If an electron has ΔE too much energy as it leaves the cavity, it will arrive back an amount $\left(\frac{\Delta E}{m_0c^2}\cdot 2\pi\right)$ late in phase and so will receive less energy than it did at the previous transit. In this way the error ΔE will be reduced and the electron will oscillate in energy and phase about the conditions corresponding to an ‘ ideal ’ electron.

The problem of stability has been considered in detail by C. Henderson *et al.* 1953 (1). Generally speaking, the energy and phase of a stable electron do not vary from that of the ideal electron by more than $\pm 0{\cdot}1$ rest masses and $\pm 10°$. The stable regions are independent of the energy of the electrons and so the beam from a microtron becomes relatively more mono-energetic as the final energy is increased.

4. GENERAL DESIGN OF A MICROTRON

A microtron consists, then, of a resonant cavity situated, *in vacuo*, to one side of a cylindrical magnetic field which is normally provided by an electromagnet so that small adjustments to the value of B can readily be made. The pressure inside the vacuum tank does not have to be extremely low ($p < 10^{-4}$ mm. of mercury is sufficient) as the path length for the electrons is comparatively short.

The resonant cavity can take various forms but a typical design consists of a flat cylindrical box in which the field along the axis has been accentuated by fitting cones to the end plates. Holes of diameter ~ 1 cm drilled along the axes of the cones allow the electrons to pass through the cavity. All sharp edges are carefully rounded off and polished, particularly in regions of high field. The cavity is tuned through a range of ± 5 MC/S by means of a thermal expansion device attached to one of the end plates. The heating is done electrically and the resulting expansion causes a small decrease in the distance between the cones and consequently of the resonant frequency. So that positive and negative adjustments can be made to the cavity frequency, the cavity is tuned on the test bench so that it is at the operating frequency when the tuning current is at half its maximum value.

The resonant cavity is fed with pulses of r.f. power a few microseconds long at a repetition rate of the order of 100 pulses per second. If the shunt impedance of the cavity at resonance is 1 megohm the power required to energize the cavity to a peak voltage of 600 KV is $\dfrac{(600 \times 10^3)^2}{2 \times 10^6} = 0{\cdot}18$ megawatt during the pulse. If this power is obtained from a self-oscillator such as a magnetron it is necessary for stability to have a resistive element in the system which takes about an equal amount of power so that ~ 0·36 megawatts are required before considering beam loading etc. which brings the power requirement up to between 0·5 and 1 MW.

The gap between the cones is usually about 0·8 cm so the peak field is about three-quarters of a million volts per cm. This field is sufficient to cause copious electron emission from the tips of the cones themselves. These electrons are suitable for acceleration and save the necessity for a separate injection system. To gain some measure of control on the emission the cones can first be highly polished and then a thin aluminium sleeve inserted into one of the cones so that its end protrudes a few thousandths of an inch into the gap of the resonator. The emission from the sleeve can be controlled by varying the depth of insertion, the roughness of its end, etc. Once the cavity starts to emit, any additional power causes a rapid increase in electron emission and little increase in voltage.

As a result of this self-stabilizing action the emission must be adjusted so that the cavity settles down at a peak voltage between 0·5 and 0·6 MV.

An additional advantage of using a sleeve in this way is that the loading on the cavity is reduced as the unwanted emission during the alternate half cycles is reduced to a minimum. It is a rough and ready method which has, however, been found to be quite satisfactory. Electron guns have been introduced into microtrons, in particular by O. Wernholm, who uses a gun pulsed at voltages up to 80 kv and situated to one side of the cavity entrance so that the issuing stream of electrons, bent into an arc of radius ~1 cm, pass into the cavity. In this way he has been able to obtain average currents of 10^{-6} amp at 5 MeV, using a duty cycle of only 3×10^{-5}.

Vertical focussing in a microtron is normally achieved by having a small region of field near the cavity which is decreasing with radius so that an electron which moves away from the median plane sees a force which has a component directing it to the median plane each time it passes through the cavity. In small microtrons the normal ‘ fall-off ’ field can be used for this purpose (fig. 3), but in larger machines a carefully shimmed field must be used to prevent undesirable precession of the orbits.

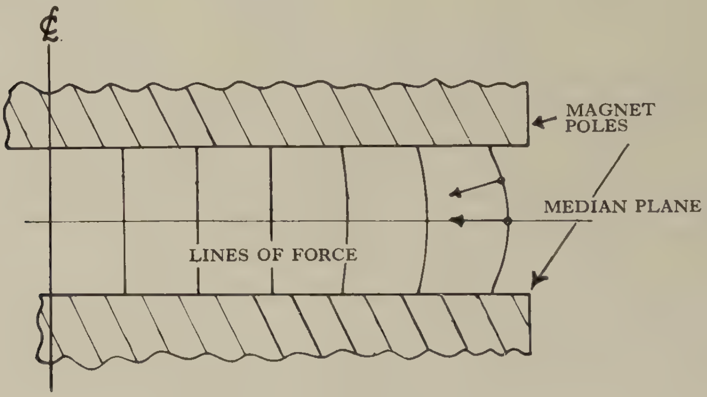


Fig. 3. Vertical focussing.

Extraction of the beam from a microtron is very simple as the orbits, in a d.c. magnetic field, are well spaced apart at just the place where the beam is to be extracted (see fig. 2). By placing an iron pipe at this point, with its axis tangential to the final orbit, the electrons pass into a field free region and so out of the machine. This iron pipe causes some disturbance of the previous orbits, but this is only serious in the case of larger machines where a symmetrical dummy extractor has to be introduced; the extractor tubes are tapered to reduce the effect on the previous orbit as much as possible. As a typical example, the efficiency of the extractor on the 4½ MeV microtron in this laboratory is ~50 per cent and the extracted beam has an energy spread of ±½ per cent and an angular spread of ±1½° horizontally and ±0·3° vertically. The output current (duty cycle 10^{-4}) is ~10^{-7} amp. average.

5. MICROTRONS FOR ENERGIES ABOVE 5 MEV

A number of microtrons have been built for energies of a few MeV and recently one has been brought into operation at 29 MeV (Aitken and Jennings 1958) and is shown in fig. 4. The final energy is determined primarily by the physical size of the magnet, the larger the magnet the larger the number of orbits, each orbit in the normal mode increasing the energy by ½ MeV. However, there are one or two points which must be carefully considered as the energy is increased.

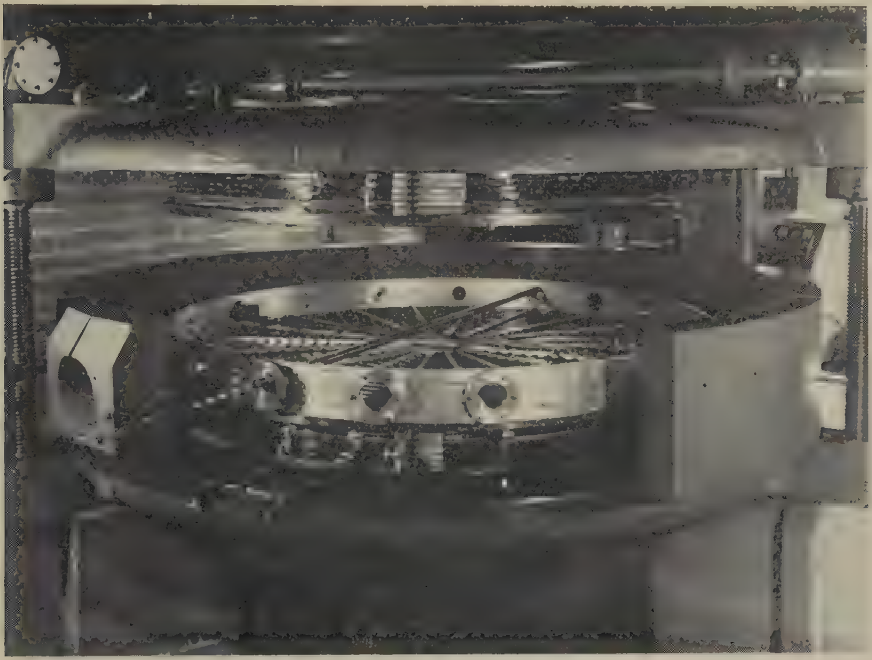


Fig. 4. 29 MeV microtron with upper pole lifted and some of the steel blocks removed.

First of all, the orbits have all to pass through a small hole, of the order of 1 cm in diameter, in the resonant cavity. At 29 MeV the final orbit has a diameter of about 2 metres and great care must be taken to make the magnetic field sufficiently uniform to prevent any appreciable precession of the orbits across the hole of the cavity. The required accuracy is achieved by careful machining and the use of pole face windings for the final adjustments. A set of concentric pole face windings are also used to correct for the small drop in field to the centre which is obtained with this type of magnet which has its return path at the outer edge of each casting. The return path consists of a ring of carefully ground blocks which also act as spacers between the upper and lower poles.

The r.f. supply is similar to that of the smaller machine, except for the use of a slightly longer pulse length (3 instead of 2 microseconds) to offset the increased acceleration time which previously was negligible and is now a little over ½ microsecond. Since the build-up time of the cavity is approximately the same, the use of a longer pulse more than doubles the output of the machine.

The poles of the 29 MeV microtron are 80 in. in diameter and the vacuum tank is formed by using a ring of aluminium which is sealed between the upper and lower poles. The upper pole can be lifted by means of built-in jacks.

The output current of this machine is $\sim 10^{-8}$ amp. average for a duty cycle of 3×10^{-4}. The extraction efficiency is not quite so high as that of the small machine. In fig. 4 can be seen a tube which passes across the vacuum tank, slightly above the median plane. This tube contains an electron resonance probe which is used to make measurements of the magnetic field and in this way, knowing the radius of the final orbit, the energy of the output beam can be found.

What are the possibilities of increasing the energy of this type of accelerator above 30 MeV? It would be extremely difficult to make a larger magnet to the required accuracy for operation in normal mode and instead it is better to consider the possibility of making use of a higher voltage across the resonant cavity and a correspondingly higher magnetic field. For electrons starting at rest the normal mode is the optimum, and nothing is gained by (say) doubling the energy gain per turn–alternate orbits would simply be omitted, the energy of the electrons increasing in the following sequence

3 5 7 9 11 etc. (total energy in units of m_0c^2)

and the time for each orbit being

3 5 7 9 11 etc. (time in periods of r.f. supply)

for the normal field. This field could not be doubled, as the time sequence would become $1\frac{1}{2}$ $2\frac{1}{2}$ $3\frac{1}{2}$ $4\frac{1}{2}$ $5\frac{1}{2}$ etc. (see conditions for resonance) However, if one could start the electrons with an initial kinetic energy equal to one rest mass, m_0c^2, then the energy sequence would become

(2) 4 6 8 10 etc. and the field could now be doubled, to give a time sequence of

(1) 2 3 4 5 etc.

This is basically the idea proposed by D. K. Aitken (1957). He suggested the use of a plane sided cavity and an emitter situated in the centre of the gap so that, with ~ 1 MV across the gap, electrons on the first transit would only gain one rest mass and so emerge with a total energy of two rest masses. On the next transit they would gain the full two rest masses and so on.

It is not necessary for the effective voltage across the gap to be exactly double that for the normal mode. If the voltage across the gap was increased to V MV ,and the electrons were started at such a point in the gap that they gained an energy of $e(V-\cdot 5)$ MeV on their first transit they would then emerge with a total energy of ev MeV and the energy sequence would then be in steps of ev MeV. The magnetic field would have to be increased by $V/0\cdot 5$ and consequently the output energy would be increased by the same ratio.

When the magnet for the 29 MeV machine was designed allowance was made for increasing the field above that required for the normal mode, and preliminary experiments on this method of increasing the output energy have been very hopeful. The main difficulty is due to the first orbit now having a much smaller radius.

6. CONCLUSION

The energy of this type of accelerator is probably limited to a maximum of about 100 MeV. In this range the machine is very suitable for research purposes such as studies on the scattering of electrons, the spectra of high energy X-rays, etc., where a primary beam having a small energy spread and angular divergence is a great advantage. While the output intensity is generally ample for such work it is not, usually sufficient for industrial and medical applications and the higher cost of a linear accelerator is then worth while. The microtron is the type of machine which can be, and to date always has been, built by the members of small research groups with outside help for only one or two of the larger items, such as the magnet poles.

I would like to finish by saying how much I have appreciated the friendly co-operation of all members of the Microtron Group at University College during the last few years.

REFERENCES

AITKEN, D. K., 1957, *Proc. Phys. Soc.* A, **70,** 550.
AITKEN, D. K. and JENNINGS, R. E., 1958, *Nature,* **181,** 1726.
HENDERSON, C., HEYMANN, F. F. and JENNINGS, R. E., 1953, *Proc. Phys. Soc.* B, **66,** 41.
HENDERSON, C., HEYMANN, F. F., and JENNINGS, R. E., 1953, *Proc. Phys. Soc.* B, **66,** 654.
KAISER, H. F., 1956, I.R.E. Transactions of the Professional Group on Nuclear Science, Volume NS-3, 17.
REDHEAD, P. A. and LECAINE, H., and HENDERSON, W. J., 1950, *Canadian Journal of Research,* A, **28,** 73.
VEKSLER, V., 1945, *Journal of Physics, U.S.S.R.,* **9,** 153.
WERNHOLM, O., Private Communication.

Additional Important References

BELL, J. S., 1953, Vertical Focusing in the Microtron, *Proc. Phys. Soc.,* B, **66,** 802–804.
MILLS, B. Y., 1950, A Million-Volt Resonant-Cavity X-ray Tube, Proc. I.E.E., **97,** III, 425

The Author:
Dr. R. E. Jennings, Ph. D., is a Lecturer in Physics at University College, London.

Electromagnetic Induction

by R. R. BIRSS
Imperial College, London

SUMMARY

Existing rules for calculating induced e.m.f. are examined and a non-relativistic expression is developed for the force on a charged particle moving in the vicinity of a time-varying but stationary source of magnetic field. This expression is used to predict the e.m.f. induced in three basic electromagnetic devices. The controversial experiment in which a long cylindrical permanent magnet is rotated about its axis of symmetry is discussed, together with a new experiment of this type which suggests that, although the field may be regarded as rotating, this does not, in itself, provide any indication as to the location of the seat of the e.m.f. in closed circuits.

---

1. INTRODUCTION

Faraday summarized the results of his experiments on magneto-electric induction only in the form of a qualitative rule. The familiar mathematical formulation

$$\text{e.m.f.} = \oint \mathbf{E} \,.\, \mathrm{d}\mathbf{l} = -\frac{\partial \Phi}{\partial t}, \tag{1}$$

where

$$\Phi = \iint \mathbf{B} \,.\, \mathrm{d}\mathbf{S},$$

was given by Neumann, a contemporary of Faraday, and was based principally on an analysis of the work of Ampère. The line integral in eqn. (1) is usually evaluated around the closed electrical circuit itself whereas Φ is the integral of $\mathbf{B}$ over any surface, $\mathbf{S}$, bounded by the path of integration.

Over one hundred years after its initial pronouncement, the interpretation of (1) is often still attended by uncertainty and controversy. This unsatisfactory situation has arisen chiefly because the flux-cutting rule (Carter 1954)

$$\mathbf{E} = \mathbf{v} \times \mathbf{B}, \tag{2}$$

gives a rival formulation but also because of the so-called electromagnetic paradoxes. There is a temptation to ignore the apparent discrepancies and to rely on direct experiment. Usually, this is the solution adopted by those who teach the basic principles of electromagnetism—for it is at an elementary level that the problem of reconciling (1) with (2) is most acute. A possible *single* expression for $\mathbf{E}$ which can be interpreted unambiguously is presented in Section 2 where the consequences of replacing a fixed path of integration by a variable one are examined. Some electromagnetic paradoxes are discussed in Section 3.

2. THE FORCE ON A CHARGED PARTICLE

2.1. *Introduction*

The difficulty of reconciling eqns. (1) and (2) may best be appreciated by considering two basic electromagnetic devices, the iron-cored transformer and the homopolar generator known as Faraday's disc. Equation (1) may be successfully applied to the iron-cored transformer but its application to Faraday's disc is complicated by the fact that there is no unique linear path around which the line integral of **E** may be calculated and hence there is no precise boundary to the circuit. Thus eqn. (1) may only be applied to a homopolar machine by the introduction of an arbitrary linear boundary along a particular path. Conversely, the flux-cutting rule may readily be used to explain the operation of Faraday's disc but not that of an iron-cored transformer, in which it may be supposed that all the flux is sensibly confined to the iron.

In the special case of a closed linear rigid circuit and a constant magnetic field produced by sources which are uninfluenced by the presence of the closed circuit, then the flux through the circuit can only change if elements of the circuit cut lines of force, so that (1) may be derived immediately from (2), **v** being the velocity of an element of the circuit relative to the field sources. Consequently, it is often maintained that (1) is a fundamental physical law whilst (2) is merely a convenient computational rule. There is, however, a strong physical reason for retaining the flux-cutting rule since it may be deduced directly from the equation

$$\mathbf{F} = q(\mathbf{E} + \mathbf{v} \times \mathbf{B}) \tag{3}$$

which gives the force on an individual charge q moving with a uniform velocity **v** relative to the (time-invariant) sources of **E** and **B**. This equation has been conclusively verified experimentally, at least for velocities considerably less than the velocity of light, and it serves to explain in more fundamental terms not only induced e.m.f's but also the forces exerted on conductors carrying current in magnetic fields. A further advantage of (3) and of the flux-cutting rule is that they predict the force on an individual charge at any point rather than the integrated value round a closed circuit. Thus (2) enables the e.m.f. between any two points in the circuit to be calculated and, when the total e.m.f. is concentrated in one element of the circuit, it predicts the seat of the e.m.f. It might therefore be thought that (1) should be employed to calculate induced e.m.f's whilst (2) should be used to provide a fundamental understanding of the physics of the process. The apparatus which, perhaps, best illustrates the fallacy of this contention is the high-energy particle accelerator known as the betatron.

In the simplest form of betatron, electrons are constrained to describe a circular path, between the poles of an electromagnet, by a radial electromagnetic force $\mathbf{v} \times \mathbf{B}$. For a constant value of **B** the trajectories of the electrons may thus be satisfactorily predicted from eqn. (3), but in the normal operation of the machine the electrons are accelerated to high energies by a tangential force produced by a steady rate of decrease of the magnetic induction, **B**. In general, each electron is therefore subject to two forces, a radial force due to the existence of **B** and a tangential force due to its diminution. The tangential force cannot be reconciled with (3)—which is therefore seen to hold only for constant **B**—but it may be deduced, albeit somewhat artificially, from (1). Although the existence of the tangential force emphasizes the inadequacy of (3), and hence of (2), the existence of the radial force—even for time-invariant **B**—does not invalidate (1), for eqn. (1) predicts only the value of the tangential component of e.m.f. integrated around the 'circuit'. Thus although (1) provides the correct value for $\oint \mathbf{E} \cdot d\mathbf{l}$ it fails to predict the total force acting on an electron. Even for a closed rigid linear circuit of conducting material (which may be referred to for convenience as a ‘wire loop’) eqn. (1) is not completely satisfactory. A material circuit is to some extent simpler for it permits a unique resolution of electromagnetic forces into e.m.f’s and ponderomotive forces but, although the circuit of integration need not necessarily coincide with the wire loop, the e.m.f’s which result from the presence of transverse electric fields (the Hall effect) cannot be readily determined. For solid bodies of conducting material, or for charged particles *in vacuo*, the applicability of eqn. (1) is, of course, even more limited except in cases which are simplified by considerations of symmetry. It is therefore desirable to investigate the possibility of developing eqn. (1) to deduce an expression for the electric field at any point (i.e. the force on an individual charge) rather than the integrated value round a closed path. Such an expression would facilitate a more complete understanding of the trajectories, *in vacuo*, of charged particles in varying magnetic fields and of magneto-electric induction in ‘solid’ bodies.

2.2. *Expressions for* **E**

In terms of the magnetic induction, **B**, eqn. (1) may be written

$$\oint \mathbf{E}\,.\,d\mathbf{l} = -\frac{\partial}{\partial t}\int\!\!\int \mathbf{B}\,.\,d\mathbf{S}\,. \tag{4}$$

Since div $\mathbf{B}=0$, it is permissible to put $\mathbf{B}=\text{curl}\,\mathbf{A}$ so that (4) may be transformed by Stokes’ theorem to give

$$\oint \mathbf{E}\,.\,d\mathbf{l} = -\frac{\partial}{\partial t}\oint \mathbf{A}\,.\,d\mathbf{l} \tag{5}$$

where the variation with time of the vector **A** and of the shape and position of the path of integration both contribute to the rate of change of $\oint \mathbf{A}\,.\,d\mathbf{l}$. It is shown in the Appendix that eqn. (5) may be written in the form

$$\oint \mathbf{E}\,.\,d\mathbf{l} = \oint\left(-\frac{\partial \mathbf{A}}{\partial t}+\mathbf{v}\times\mathbf{B}\right).\,d\mathbf{l}, \tag{6}$$

where **v** is the velocity of any point on the path of integration. Equation (6), when written in the alternative form

$$\oint \mathbf{E}\,.\,d\mathbf{l} = -\frac{\partial \Phi}{\partial t}+\oint(\mathbf{v}\times\mathbf{B})\,.\,d\mathbf{l}, \tag{7}$$

illustrates a separation of the e.m.f. into a variational or transformer term and a motional or flux-cutting term. It should be emphasized, however, that this separation is only a unique one relative to the coordinate system to which the velocities, magnetic inductions and the position and configuration of the circuit are referred.

To deduce, from (6), the value of **E** at any point, it must be remembered that any acyclic function may be added to the integrands without affecting the value of the line integrals around any closed path. **E** may therefore be written in the form

$$\mathbf{E} = -\frac{\partial \mathbf{A}}{\partial t}+\mathbf{v}\times\mathbf{B}-\text{grad}\,(\phi+\psi) \tag{8}$$

where ϕ is the electrostatic potential, which exists even in the absence of changes in the magnetic field and the circuit, and ψ is an unknown scalar function. Similarly, the function **A** is not completely defined by the equation $\mathbf{B}=\text{curl}\,\mathbf{A}$, since any acyclic function may be added to **A** without altering the value of the curl. However, in view of the occurrence of the unknown function ψ, no additional restriction is imposed by postulating that div $\mathbf{A}=0$ so that $\mathbf{A}$ is thereby completely defined and may be identified with the Maxwell vector potential.

2.3. *Three basic electromagnetic devices*

Equation (8) may be readily used to explain the operation of the three electromagnetic devices discussed in sub-section 2.1 provided it is remembered that $\mathbf{E}$ is the force on an individual charge and that $\mathbf{v}$ must be interpreted as the velocity of an individual charge relative to a set of coordinate axes in which the sources of $\mathbf{B}$ are stationary. If this set of axes is stationary in the laboratory frame of reference, then the operation of the betatron shows that the term in ψ is not present, since the radial and tengential forces on each electron are observed to be given exactly by $\mathbf{v}\times\mathbf{B}$ and $-\partial\mathbf{A}/\partial t$ respectively.

In an iron-cored transformer, e.m.f's may arise from both the $\mathbf{v}\times\mathbf{B}$ and the $-\partial\mathbf{A}/\partial t$ terms. For example, consider a set of cylindrical polar coordinates, r, θ, z, in which a long cylindrical iron core, of radius a, has the z-axis as its axis of symmetry. The vector potential, $\mathbf{A}$, is everywhere perpendicular to both the radius r and the z-axis. In terms of the alternating flux density $\mathbf{B}_i$ in the iron core, the magnitude of $\mathbf{A}$ is given by $A_\theta=\frac{1}{2}rB_i$ for $r\leqslant a$ (i.e. within the core) and $A_\theta=\frac{1}{2}a^2B_i/r$ for $r\geqslant a$. The e.m.f. developed across the terminals of a single-turn coil of radius a', concentric with the core, is thus

$$-2\pi a'\frac{\partial\mathbf{A}_\theta}{\partial t}=-\pi a^2\frac{\partial B_i}{\partial t}=-\frac{\partial\Phi_i}{\partial t},$$

where Φ_i is the magnetic flux in the iron. When a current flows in this coil an e.m.f. is developed across the turn and perpendicular to the z-axis: its magnitude is $\mathbf{v}\times\mathbf{B}_0$, where $\mathbf{v}$ is the velocity possessed by individual charges by virtue of the current flowing in the coil and $\mathbf{B}_0$ is the small leakage flux density outside the iron at a'. This e.m.f. gives rise to a compensating transverse separation of static charge and hence to the Hall effect.

Since Faraday's disc operates in a constant magnetic field, the observed e.m.f. arises only from the $\mathbf{v}\times\mathbf{B}$ term. The use of eqn. (8) rather than (6) or (7) obviates the difficulty, already mentioned, that there is no precise boundary to the circuit and hence no unique linear path of integration. It shows, however, since $\mathbf{v}$ is the velocity of any point on the path of integration, that, if (6) or (7) is to be used, the motional term should be calculated by considering the path of integration to be completed by a radius vector rotating with the disc and by a line (of constantly increasing length) drawn around the periphery of the disc from this radius vector to the sliding contact. The variational term,

$$-\frac{\partial\Phi}{\partial t}=-\oint\frac{\partial\mathbf{A}}{\partial t}\cdot d\mathbf{l}, \tag{9}$$

is, of course, zero since, as indicated by the partial derivatives, it is evaluated by considering the path of integration to be stationary. It may be mentioned at this point that the change in length of the peripheral segment of the path of integration need cause no difficulty since the derivation of (6) and (7) applies to a circuit which is being deformed both by flexure and extension. Nor does the mere existence of the sliding contact cause any difficulty *per se*.

3. SOME ELECTROMAGNETIC PARADOXES

3.1. *Stationary magnetic-field sources*

It was stated in section 2.2. that eqn. (8) gives the force on an individual charge if $\mathbf{v}$ is interpreted as the velocity of the charge relative to a set of co-ordinate axes in which the sources of $\mathbf{B}$ are stationary, and it is thereby assumed that none of the sources of $\mathbf{B}$ move relative to each other. If this is true, and if the sources of $\mathbf{B}$ are also stationary in the laboratory frame of reference, the force on an individual charge may be written in the form

$$\mathbf{E} = -\frac{\partial \mathbf{A}}{\partial t} + \mathbf{v} \times \mathbf{B}, \tag{10}$$

where, for simplicity, the electrostatic term $-\operatorname{grad}\phi$ has been omitted. Under such circumstances the only paradoxes which can arise are those which result from employing (1) or (2) or both instead of (10). A typical example, which also illustrates the danger of attaching too much importance to the concept of flux-linkages, is afforded by the experiments of Blondel (1914). In these experiments a coil of wire is unwound in a steady magnetic field so that the number of turns linking magnetic flux is progressively decreased. Despite the fact that the flux-linkages are progressively decreased, no e.m.f. is generated unless the connection, from the axis of the coil to the last turn, rotates with the coil. This result is, of course, to be expected from eqn. (10), although believers in the dictum that the e.m.f. is equal to the rate of diminution of flux-linkages may draw some comfort from the fact that it is only when a ‘ solid ’ body is used to complete the circuit or when the number of ‘ turns ’ changes progressively but discontinuously that the expected e.m.f. is not obtained. Nor are these the only *bêtes noires*, for paradoxical results can also arise from a procedure known as ‘ substitution of circuit ’. A typical example is that of an ordinary tubular sliding resistor situated in a magnetic field. Again the number of flux-linkages can be changed without inducing an e.m.f. but only by substitution of circuit. A more satisfactory explanation is provided by eqn. (10).

3.2. *Moving magnetic-field sources*

If, relative to the laboratory frame of reference, all the sources of $\mathbf{B}$ are moving with the same constant (linear) velocity, $\mathbf{u}$, which is small compared with the velocity of light, then the force on an individual charge is

$$\mathbf{E} = -\frac{\partial \mathbf{A}}{\partial t} + (\mathbf{v} - \mathbf{u}) \times \mathbf{B}, \tag{11}$$

where $\mathbf{v}$ is the velocity of the charge in the laboratory frame of reference and $\partial \mathbf{A}/\partial t$ is calculated relative to axes which move with the sources of $\mathbf{B}$. An interesting paradox which arises with a linearly moving magnet is Herring's experiment (1908) in which a magnet is forced between a metal spring clip, the two halves of which are joined electrically only by an external circuit. From eqn. (11) it may be seen that, since $\partial \mathbf{A}/\partial t = 0$, the value of $\mathbf{E}$ at any point depends only on the relative velocity of the circuit and the magnet, and the e.m.f. obtained is therefore the same as if the clip and circuit were in motion and the magnet at rest. Integration of $\mathbf{E}$ from one side of the clip to the other around the circuit gives no resultant e.m.f. (This may be most easily seen for the special case in which the clip is at the extreme end of the magnet. If the e.m.f. induced in a circuit in which the clip is bridged by a wire sliding over the end face of the magnet is compared with the rate of change of flux through that circuit it may be seen that the net e.m.f. around the circuit bridged by the magnet is zero.)

If two or more sources of **B** are moving with different velocities then a number of terms of the same form as (11) must be used, summed over all the sources of **B**. For example if two similar magnetic poles are symmetrically situated near a conductor so that instantaneously they lie on a straight line through the conductor and perpendicular to it, whilst moving at right angles to this line and to the conductor in opposite directions, then the e.m.f. obtained is double that appropriate to a single pole although the conductor itself is in zero magnetic field. This result is often stated to be paradoxical. Similarly a summation of eqn. (11) may be applied to any number of linearly moving sources of **B** although its application is seldom straightforward when materials are present which act as magnetic shields without themselves being primary sources of **B**.

3.3. *Rotating magnetic-field sources*

More complicated—and more interesting—cases arise when the sources of **B** are rotating in the laboratory frame of reference; for example, a long cylindrical bar magnet rotating about its axis. Again the $-\partial \mathbf{A}/\partial t$ term in eqn. (8) is zero and a stationary charge should experience a force from the $\mathbf{v} \times \mathbf{B}$ term or, more exactly, from a $\Sigma \mathbf{v} \times \mathbf{B}$ term where the summation extends over the elementary dipoles forming the magnet (each moving with a linear instantaneous velocity $\mathbf{v}$). However, there is no experimental evidence, as there is with a stationary magnet, that the term $-\text{grad}\,\psi$ is zero and there is thus the possibility that such a term may be introduced by the rotation of the magnet. It is not therefore clear whether or not a stationary charge in the vicinity of the magnet will be acted upon by a force. For example, in the modified Faraday's disc arrangement shown in fig. 1, when the disc D spins in the field of the magnet M a current flows through the circuit AG_1B which is completed by means of sliding contacts on the periphery and the axle of the disc. If, alternatively, the disc is stationary whilst the magnet rotates about the common axis EB no current is registered by G_1. However, this does not prove that $\mathbf{E}$ is everywhere zero; for $\Sigma \mathbf{v} \times \mathbf{B} - \text{grad}\,\psi$ integrates to zero around the complete circuit whatever the value of ψ. Further, although a galvanometer G_2 connected to brushes E and F bearing on the magnet as indicated shows a deflection, this could be interpreted as being due

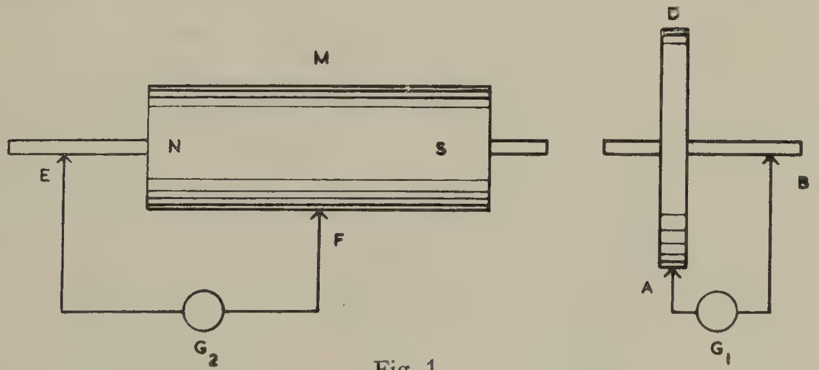


Fig. 1.

either to a field of force (**E**) surrounding the magnet or to one within the magnet itself. These difficulties most usually present themselves in terms of a conceived motion of the magnetic field surrounding the magnet. This point of view originated with Faraday, who was accustomed to thinking in terms of lines or tubes of magnetic force and who, in his Diary of July 14, 1851, asked:

> " 11345 When the magnet is still and the wire is moving, it seems unlikely that the current should be generated anywhere else than in the moving wire; for its motion or quiescence makes all the difference. But then, when the magnet is moving, where is the current then generated? In the wire across which the curves, that may be supposed to move with the magnet, are passing? Or in the magnet, which may be supposed to be moving (as the wire did) whilst the curves are considered still? Do the lines of force revolve with the magnet or do they not? "

If a motional e.m.f. is conceived to arise from motion relative to lines of magnetic force rather than motion relative to the field sources themselves, then the concept is only a useful one in so far as the motion of the lines of force may be determined for a given motion of the field sources. The connection between eqn. (8) and the concept of motion relative to lines of force is an interesting one. If a magnetic-field source is stationary in the laboratory frame of reference, the lines of magnetic force must also be assumed to be stationary since it is known from experiment that $\psi=0$. If the field source moves with a constant velocity $\mathbf{v}$ which is small compared with the velocity of light, then, because of the equivalence of inertial frames of reference, it may be supposed that ψ is again zero and hence the lines of force must be assumed to move with the field sources. A rotating field source, however, is not stationary in an inertial frame of reference and therefore the principle of equivalence cannot be invoked to deduce that ψ must be zero. A finite value of ψ does not, of course, alter the e.m.f. induced in a closed circuit, for $-\operatorname{grad}\psi$ integrates to zero around any closed path, but it may affect the location of the seat of the e.m.f.

For example, in the arrangement shown in fig. 1, a rotation of the magnet M results in a deflection of the galvanometer G_2 which, if $\psi=0$, arises wholly in the circuit EG_2F. If $\psi\neq 0$ the observed e.m.f. is no longer generated wholly in this circuit but partly in the magnet itself and, for a certain finite value of ψ, the seat of the e.m.f. is wholly in the magnet. Thus, if the e.m.f. is conceived to arise from motion relative to lines of force, the lines of force may be assumed to be either stationary or rotating. Both assumptions are widely held although, as yet, the hypothesis that the lines of force rotate at a speed which is different from that at which the magnet rotates has found no adherents.

Similar considerations arise for the stationary circuit AG_1B of fig. 1, in which no e.m.f. is generated by a rotation of the magnet M. Although it may be argued that it is inconsistent to concede a variation in the place where the e.m.f. is generated since the relative motion of the magnet and the disc is the same as in the conventional Faraday's disc experiment, it must again be remembered that rotation, unlike translation, is absolute and not relative.

3.4. *A meaningless question*?

The quotation from Faraday's Diary given above poses two questions, viz.

(*a*) When the magnet is rotated where is the e.m.f. generated?

(*b*) Do the lines of force rotate with the magnet or do they not?

It is obvious that Faraday considered these questions as being two alternative but equivalent formulations of the same problem. This view was shared by Steinmetz, Barnett (1912, 1913), Kennard (1913, 1916) and Pegram (1917) who also believed, with Faraday, that the lines of force do not rotate and that an e.m.f. is induced in the rotating magnet. Weber—to whom the term unipolar induction is due—believed that they did rotate, a view which was shared by Cullwick (1938), Sir Oliver Lodge (although this has been subsequently disputed), Andrew Gray, Cramp (1933) and Cramp and Norgrove (1936). A recent paper by Hammond (1954) promoted further discussion; Professor Cullwick and Dr. Wilkinson maintained that the e.m.f. was generated in the stationary leads whereas Hammond said that he still inclined to the view that the seat of the e.m.f. was in the rotating magnet.

Whatever the true state of affairs, it had long been known (Preston 1885) that the integrated value of **E** around the complete circuit was the same under either hypothesis and it was not therefore expected that experiments on complete circuits would settle the issue. In such circumstances it was not long before it was maintained that the question of whether the field rotates with the magnet or not is fundamentally a meaningless one. This point of view was put forward by Poincaré (1900), Abraham, Kennard and Föppl and found its most recent and most vehement protagonist in Professor G. W. O. Howe (1934, 1936 a, b, 1950).

The most cogent expressions of the belief that the question is in reality meaningless are due to Howe. He pointed out that it would be meaningless to ask whether the temperature field of a hot body rotated with it or whether the uniform distribution of illumination from a lamp rotated with it. Further, he maintained that the concept of a rotating field implies that the lines of force have an individuality, that one can be distinguished from the next: but lines of force are mathematical abstractions, their density is decided only by convention,

> " One is not justified in regarding them as physical entities like bristles that one can chalk-mark and thus trace their movement ".

Against attaching too much physical significance to lines of force Howe argued that although, under certain circumstances, e.m.f.'s may be generated, and the magnitudes of these e.m.f.'s may be *proportional* to the number of lines of force cut, this does not mean that the e.m.f.'s are *due to* the cutting of physical lines. These, and similar, arguments have been advanced so frequently and with such persistence that it is of considerable interest to consider them in more detail.

The analogy between a magnetic field and, for example, a temperature field is a doubtful one since it is not possible to deal purely with a magnetic field isolated from associated electric effects. The possibility of an electrical effect associated with the rotation of a magnet is neglected in the analogy with a temperature field, yet such an effect can be made a criterion of rotation. The claim that the concept of a rotating field implies the individuality or distinguishability of lines of force overlooks the elementary distinction between a necessary and a sufficient condition. Following the path of an individual line of force might well be sufficient to define the rotation of a magnetic field; the impossibility of this procedure does not, however, render invalid the concept of field rotation since alternative criteria may exist which permit a unique determination of whether the field is rotating or not. Furthermore, such criteria need depend in no way on the possibility of tracing the movement of individual lines of force. Such a criterion would, for example, be to define the rotation of a field in terms of the force on a stationary charge. The argument against attaching too much physical significance to lines of force cannot be treated too seriously for, as with many other physical concepts, their justification is only on the grounds of convenience. Whether the cutting of lines and the e.m.f. are cause and effect or not they certainly co-exist and mathematically this is the same thing. There is much point in Professor Cramp's argument (1938) that relative motion between a conductor and the lines of force confers physical reality to the lines, in other words a physical explanation of the e.m.f. involves the physical existence of the lines of force. Relative motion between a conductor and a structureless medium is a 'contradiction in terms—an abrogation of essential consistency', it is 'operationally meaningless'.

As well as advancing counter-arguments to the charge that it is a meaningless question to ask whether the field rotates with the magnet or not, it is also possible to perform an experiment which proves that the question is not meaningless and which also has some bearing on whether the field should, in fact, be regarded as rotating or not. This experiment is discussed in section 3.5. It is therefore desirable to consider whether questions (*a*) and (*b*) are, in reality, alternative formulations of the same question. The location of the seat of the e.m.f. is a question of fact which, unfortunately, cannot be answered by experiments on closed circuits. Measuring the force on an isolated charged particle would, of course, be sufficient to answer question (*a*). Such experiments—for example rotation of the magnetic poles in the betatron experiment described in section 2.1.—have, however, not yet been performed. It may therefore be concluded that (*a*) is a question of fact which has not yet been decided by experiment but to which a definite answer exists. Question (*b*) is not a question of fact but of convention: if a motional e.m.f. is conceived to arise from motion relative to lines of magnetic force then a set of rules which governs the motion of the lines of force must be formulated so as not to contradict any existing experimental results. It follows that an answer to (*a*) in conjunction with all the other available experimental evidence, must either automatically answer (*b*) or invalidate altogether the concept of motion relative to lines of force. It is, however, possible to devise an experiment which will provide an answer to (*b*) that must be regarded as valid until it is verified or invalidated by a decision on (*a*), or until it is invalidated by further experiments to decide (*b*). Such an experiment is described in section 3.5 but it should be emphasized that a decision as to whether or not the field may be regarded as rotating does not, in itself, provide any indication as to the location of the seat of the e.m.f.

3.5. *A new experiment*

An experiment has recently been performed with the apparatus illustrated diagrammatically in fig. 2. M and M′ are two hollow cylinders of Alcomax magnetized with adjacent surfaces of opposite polarity. Y is a soft iron yoke, of square cross section. D and D′ are copper discs, A is an annular brass disc insulated from D′C which is a copper rod. The two systems (D′DC) and (AB, MYM′) can both be rotated independently of each other by means of a motor and a driving arrangement on the end C (which have been omitted for clarity). This driving system is surrounded by an electrostatic screen and it is not in electrical contact with the rest of the apparatus. The wire AB is insulated from the yoke Y and the electrical circuit, which is completed by the galvanometer G, is D′DBA.

When the disc rotates and the yoke is stationary it is found that an e.m.f. is obtained proportional to the speed of rotation, whilst with the yoke rotating and the disc stationary an equal and opposite e.m.f. is obtained. When both disc and yoke rotate together no galvanometer deflection is observed. It may be assumed that the magnetic field is confined to the yoke and the pole-pieces MM′, except in the region of the air gap. A quantitative analysis of the experimental data confirmed that the effect of the stray field could, in fact, be neglected. It may be noted that the results obtained with this apparatus are in complete agreement with eqn. (8) because the $-\partial \mathbf{A}/\partial t$ term vanishes since it is calculated relative to coordinate axes which rotate with the yoke and pole-pieces.

If it is asked what bearing the results of this experiment have on question (*b*) of section 3.4, a plausible answer would be to say that, since an e.m.f. can be induced when the disc D is stationary in a magnetic field which is undergoing no change either in magnitude or direction and when the connecting circuit AB is rotating in zero magnetic field, then the magnetic field between the pole-pieces M and M′ must be regarded as rotating. However, since it could be argued that this e.m.f. is a variational one which arises from a change in the flux linking the complete circuit, it is necessary to interpret the experimental data with reference to the rules which must be complied with if observed electromagnetic effects are to be correctly explained by a consideration of lines of magnetic force.

These rules have been given by Bewley (1952) and may be explained by reference to eqn. (7). They state that, in a closed circuit, the motional e.m.f. is to be calculated by interpreting $\mathbf{v}$ as the velocity of any point of the circuit relative to the lines of force and by assuming that $\mathbf{v}$ and $\mathbf{B}$ are instantaneously constant, whilst the variational e.m.f. is to be calculated by assuming that the circuit is instantaneously stationary relative to the lines of force and that only $\mathbf{B}$ varies. For a closed *linear* circuit these rules are ambiguous only in so far as the motion of the lines of force is uncertain, but the introduction of a pluridimensional body complicates the evaluation of both motional and variational components since the circuit (or path of integration) need not necessarily be the same for both components. However, a certain simplification is afforded by the fact that, if two particular paths of integration are postulated, these paths must depend only on the relative disposition and motion of the various parts of the actual circuit and in no way on whether or not the magnet is rotating.

For example, however the circuit D′DBA of fig. 2 is considered to be completed through the disc D, it must be completed in the same way as is the circuit AG_1B of fig. 1. When either the circuit AG_1B or the disc D rotates, the observed e.m.f. is a motional one whether or not the magnet M rotates and whether or not, if it does, the field rotates with it. The $\oint(\mathbf{v}\times\mathbf{B})\,.\,d\mathbf{l}$ term must therefore be calculated around a path of integration which includes, as described in section 2.3, a radius vector rotating with the disc and a peripheral segment of constantly increasing length. Since the variational component is absent, the $-\partial\Phi/\partial t$ term must be calculated relative to a circuit which is completed by a straight line from A to the axis of D and which thus does not link any magnetic flux. Making the same assumptions for the apparatus shown in fig. 2, it follows that the experimental results described above may only be interpreted by assuming that the lines of force rotate with the magnet. This does not, in itself, provide any indication as to the location of the seat of the e.m.f.

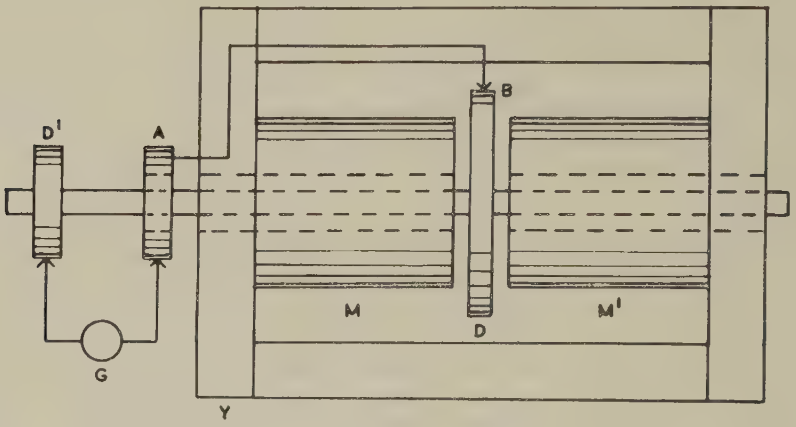


Fig. 2.

It may, of course, be objected that proving the field rotates with the unsymmetrical magnet shown in fig. 2 does not prove that it would rotate with a completely symmetrical magnet—nor indeed does it. However, even for a symmetrical rotating magnet, relativistic electromagnetic theory (Cullwick 1954) confirms that the field may be regarded as rotating and assigns the seat of the e.m.f. to the stationary conductors. To pursue this further would, however, be beyond the scope of the present article for the treatment is controversial and one in which the permitted limits of scientific heterodoxy are soon reached.

Acknowledgment

The experiment described in section 3.5. was performed while the author was a member of the Physics Department of Nottingham University and he is indebted to Professor L. F. Bates for providing the facilities for the construction of the apparatus and for the experimental work.

References

Barnett, S. J., 1912, *Phys. Rev.*, **35**, 324; 1913, *Ibid.*, **2**, 323.

Bewley, L. V., 1952, *Flux Linkages and Electromagnetic Induction*, p. 26 (New York: MacMillan).

Blondel, A., 1914, *Comptes Rendus*, **1959**, 674.

Carter, G. W., 1954, *The Electromagnetic Field*, p. 149 (London: Longmans).

Cramp, W., 1933, *Electrician*, **111**, 471; 1938, *Ibid.*, **121**, 371.

Cramp, W., and Norgrove, E. H., 1936, *Journal I.E.E.*, **78**, 481.

Cullwick, E. G., 1938, *Electrician*, **121**, 286 and 467; 1954, *Proc. I.E.E.* (Part I), **101**, 162.

Hammond, P., 1954, *Proc. I.E.E.* (Part I), **101**, 147.

Herring, C., 1908, *Trans. Amer. I.E.E.*, **27**, 1341.

Howe, G. W. O., 1934, *Electrician*, **113**, 329; 1936 a, *Ibid.*, **117**, 191, 493 and 625; 1936 b, *Journal I.E.E.*, **79**, 345; 1950, *Proc. I.E.E.* (Part I), **97**, 129.
Kennard, E. H., 1913, *Phys. Rev.*, **1**, 355; 1916, *Ibid.*, **7**, 399.
Pegram, G. B., 1917, *Phys. Rev.*, **10**, 591.
Poincaré, H., 1900, *L'Éclairage Électrique*, **23**, 46.
Preston, S. T., 1885, *Phil. Mag.*, **19**, 131 and 215.

APPENDIX

If the shape and position of the path of integration do not change with time, then $\frac{\partial}{\partial t}\oint \mathbf{A}\,.\,d\mathbf{l} = \oint \frac{\partial \mathbf{A}}{\partial t}\,.\,d\mathbf{l}$. In general, however, an additional term is present due to variations in the shape and position of the path of integration. This term may be evaluated by considering a finite section PQ of the path of integration which moves to the position P′Q′ after a short interval of time dt. If $\mathbf{A}$ is considered to be constant over the interval dt then the change due to the motion of the curve alone is

$$\int_{P'}^{Q'} \mathbf{A}\,.\,d\mathbf{l} - \int_{P}^{Q} \mathbf{A}\,.\,d\mathbf{l} = \int_{Q}^{Q'} \mathbf{A}\,.\,d\mathbf{l} - \int_{P}^{P'} \mathbf{A}\,.\,d\mathbf{l} - \oint^{PQQ'P'} \mathbf{A}\,.\,d\mathbf{l}\,.$$

If $\mathbf{v}$ is the velocity of any point on the path of integration, the first two terms on the right hand side of this equation represent dt times the difference between the values of $\mathbf{A}\,.\,\mathbf{v}$ at Q and at P, whilst the third term may be expressed, by Stokes' theorem, as a surface integral over elements of area $dt(d\mathbf{l}\times\mathbf{v})$. Thus

$$\int_{P'}^{Q'} \mathbf{A}\,.\,d\mathbf{l} - \int_{P}^{Q} \mathbf{A}\,.\,d\mathbf{l} = dt\int_{P}^{Q} d\mathbf{l}\,.\,\text{grad}\,(\mathbf{A}\,.\,\mathbf{v}) - dt\int_{P}^{Q} (\mathbf{v}\times\text{curl}\,\mathbf{A})\,.\,d\mathbf{l}\,.$$

The complete rate of change of $\int_{P}^{Q} \mathbf{A}\,.\,d\mathbf{l}$ is thus

$\frac{\partial}{\partial t}\int_{P}^{Q} \mathbf{A}\,.\,d\mathbf{l} = \int_{P}^{Q}\left[\frac{\partial \mathbf{A}}{\partial t} + \text{grad}\,(\mathbf{A}\,.\,\mathbf{v}) - \mathbf{v}\times\text{curl}\,\mathbf{A}\right].\,d\mathbf{l}$ and eqn. (5) may therefore be written in the form $\oint \mathbf{E}\,.\,d\mathbf{l} = \oint\left[-\frac{\partial \mathbf{A}}{\partial t} + \mathbf{v}\times\mathbf{B}\right].\,d\mathbf{l}\,.$

The Author :

Dr. R. R. Birss is at present engaged at Imperial College upon research in solid state physics with particular emphasis on ferromagnetic materials. He is also interested in special relativity and the electrodynamics of moving bodies.

Physical Science in the Training College

by E. J. WENHAM

City of Worcester Training College

Dr. Lester's interesting account of the stimulating project being undertaken at Hatfield Technical College in conjunction with Balls Park Teacher Training College stresses that the aim of the course described is not to produce ' science teachers ' but to produce ' science-conscious teachers '. That this aim has been achieved cannot be doubted, for all six girls involved in the first course accepted posts as teachers of science in secondary modern schools. There the experienced teachers in charge of the subject will welcome them as valuable reinforcements.

In general, these experienced teachers have received their training of two, and, sometimes of three, years duration in a general Training College. There they have undertaken courses in one or two ' main ' subjects, and in Education, together with shorter courses in certain other subjects (often called ' curriculum courses '). In the field of science, most Colleges have had some course (by whatever name) in Biology, many in a ' General Science ' and a few in the Physical Sciences. For some years now, this has been regarded as far from satisfactory and the Colleges and the Ministry have been considering the problem of providing teachers competent to take charge, both of the subject of science and of the laboratories in the secondary modern schools. This is but one facet of the general problem posed by the well-recognized shortage of science and mathematics teachers in all schools. During the years 1956 and 1957, the Minister of Education took concrete steps to remedy this particular shortage by encouraging certain Training Colleges to open a number of new science courses for serving teachers—additions to the Supplementary Courses then available in a few Colleges. The Colleges chosen were those with reasonably adequate facilities in the shape of laboratories and staff, where science courses of one sort or another were already in existence. Thus, in the City of Worcester Training College, which can perhaps be regarded as typical of this group, well-established ' main ' courses in Physical Science, Biological Science, General Science and Rural Science provided a base upon which was built a Supplementary Course for about a dozen serving teachers (or students who had just completed a suitable College course of two years duration). The new course, which lasted for one year, aimed at a general raising of the student's knowledge of the sciences, with particular reference to the needs of the schools.

Within a year, it became obligatory for those incoming students of this College, who wished to follow a career as specialist science teachers, to opt to stay for this third year course, having studied the Physical and the Biological Sciences as their ' main ' subjects for the two previous years. In effect, the three-year course of teacher-training in science had begun. In autumn 1960, these students entered upon the third year of their course to devote the whole of their time to studies of science and of science teaching, having already spent at least 8 hours a week in their first year and 12 in their second year to work in the sciences. In addition they will have followed the customary courses in Education as well as the curriculum courses in Mathematics and English. In a few other colleges this pattern has already been established for several years with most encouraging results.

The autumn of 1960 also saw the commencement of the three-year course for all Training College students. This is a revolution in the pattern of teacher training in England and Wales. The projected courses in the sciences have gained much from the experience provided by the ' Supplementary Courses '. Here a curious nomenclature must be noted in that Colleges possessing the facilities required by the Supplementary Courses are deemed to possess ' Science Wings '. In the sixteen Colleges so designated, specialist courses to a high standard are being developed. All have devised patterns of work which differ in detail, as befits their positions as constituent Colleges of some eleven different Institutes of Education. At Worcester (University of Birmingham Institute of Education) students entering in September 1960, who plan to become specialist teachers of science, will spend six hours a week on Physical Science and a further six hours on Biological Science for their first year: in the second year they will add to this time allocation two hours of Handicraft (in wood, glass and metal) and two additional hours of Chemistry: in the third year they will specialize on either Physical or Biological Science for eight hours of the week, but will have six further hours available for other studies in the sciences. In addition, four hours weekly will be devoted to Education throughout the three years of the course, while in the first year certain professional studies will also be followed.

To undertake such a commitment the new College building provides two Physics laboratories, an Optics laboratory, a Chemistry laboratory, two Biology laboratories, a General Science laboratory, a lecture theatre, a workshop, a photographic dark-room, one large and four small tutorial rooms together with preparation and store rooms for each laboratory. The staff, now comprising four specialists (two Biologists, one Chemist and one Physicist) will expand to meet the demand.

The syllabus in Physical Science for these students takes the following form:

Introductory: The three fundamental concepts of time, space and matter. Their measurement, with particular reference to measurements of the solar system.

The elements of dynamics: Motion in a straight line and a circle. Newton's laws and concepts of force, momentum and energy and of angular momentum, centripetal force and moment of inertia, all with particular reference to the solar system and including a study of artificial satellites. The conservation of mass, momentum and energy. A simple treatment of relativity.

The nature and properties of matter: Chemical reaction: the development of the concepts of atoms and molecules and of the electrical nature of matter. The nature of solids, both crystalline and non-crystalline and their mechanical properties. The nature of liquids and such properties as those of surface tension and viscosity. The nature of gases and their properties—the gas laws. Simple classical kinetic theory. The relationship between the states of matter.

Heat: The concepts of heat and of temperature. The properties and effects of heat (including expansion, transfer, specific heat and change of state). Heat as a form of energy. Prime movers, and energy conversions. Fuels.

Optics: Propagation, reflection and refraction of light. Optical instruments. Colour. Photometry.

Electricity: The simple circuit and the effects of the electric current, including electrolysis. Current, potential difference and e.m.f. and their measurement. Ohm's Law and resistance. Kirchhoff's laws. Transient and alternating currents. The capacitor and the inductor. Power in the a.c. circuit.

The phenomena associated with fields: gravitational, magnetic and electric. Magnetic fields of steady currents: Ampere's Law. Electromagnetic induction: Faraday's Law. Electromagnetic machinery. The magnetic properties of matter.

Periodic phenomena: Simple harmonic motion, oscillations of the simple pendulum and of electrical circuits. Acoustics. Longitudinal and transverse waves. The wave theory of light and the phenomena of interference, diffraction and polarization. Electromagnetic waves.

Quantum phenomena: The evidence for the existence of the fundamental particles. Simple electronics. Spectra, the Bohr atom and the Periodic Table. Conduction in semi-conductors. The photo-electric effect. Nuclear reactions—radioactivity, fission, fusion, the production of energy in the sun.

History of Physical Science: A detailed study of one or more case histories of important advances will be made in addition to a survey of the general development since the scientific revolution of the 17th century.

Teaching methods, Laboratory management and techniques.

Ancillary courses in Handicraft, Chemistry and Mathematics will be provided, whilst there will be opportunity to attend short courses in Statistics, Meteorology, Astronomy and Photography.

The treatment of the subject matter is both practical and mathematical and the students are encouraged to take their studies to as high a level in certain selected fields as is possible for them. Perhaps one might venture the hope that successful completion of such a course would bring the student to within a year of the standard of Part II of the General B.Sc. degree of the University of London. Whether or not this is so in all cases, such a course provides a more than nodding acquaintance with those aspects of scientific theory and practice which are so much in the public eye as well as a thorough appreciation of the fundamentals upon which modern science rests. And all teachers know that a really thorough knowledge of these is essential if their teaching is to be both competent and confident.

The Colleges involved in this expansion—both in numbers and in depth of study—have experienced major changes. In many cases new buildings are being erected which will provide facilities far more extensive than anything known in the past. Where suitable applicants come forward (and it is precisely here that the Colleges are seeing the vicious circle of shortage in action) new members of staff have been appointed.

Meanwhile, other Colleges, apart from these sixteen with ' wings ', have expanded their teaching of the sciences. In all, some 43 others now provide courses in General Science, 15 in Chemistry, 24 in Physics and 13 in Chemistry and Physics. As already stated, almost all provide a course in some form of Biological Science. If the hopes of the Ministry of Education, as described in the pamphlet " The Training of Teachers " are to be fulfilled, these courses will provide the schools with teachers competent to support their specialist colleagues who are drawn from the University Training Departments and the Wing Colleges, and with others who can perhaps provide some degree of specialization in the Junior School.

Whilst the development of the student's own knowledge of science is essential it is no less important that he must be in a position to see his teaching in relation to the needs and the abilities of the children he is to teach. This problem, too, the Training Colleges are attempting to tackle. So time is devoted to the consideration and the practice of the method of science teaching (which is essentially scientific): to the care and maintenance of equipment: to the development of those skills in the handling of materials which enable the teacher to improvise apparatus where necessary and to guide laboratory assistants where these are available. School syllabuses must be discussed and subject to critical analysis for the students are, by and large, drawn from Grammar Schools and so only know Grammar School science, which may have little relevance for the child in the Secondary Modern School.

The Training Colleges, then, with the encouragement of the Ministry of Education, have not neglected the challenges posed by the shortage of science teachers and by the inadequacy of much of the science teaching. So far as has been possible, they have endeavoured to produce more and better trained teachers, providing at one and the same time an adequate academic background and a degree of professional competence. Nevertheless the problem cannot be solved by this present expansion. Further development within the Colleges is an urgent necessity. The Science Masters' Association's enquiry into " Science Teaching in Secondary Modern Schools ", which is based upon replies received from some 1500 schools of the 3690 known to exist, shows that in these schools the teachers of science include:

	%
Graduates in science	18
Others who have taken a full-time supplementary course of a year's duration, or who have completed a science degree course	11
Others who have studied one or more science subjects to G.C.E. ' A ' level or its equivalent	53
Others	18

Furthermore, only 34 per cent of schools regarded themselves as adequately staffed for science teaching.

To the question, " What is the most important need to improve science teaching in your school ? " 65 per cent of Head Teachers replied that more science teachers were needed. " It follows that, in the country as a whole, no fewer than 2,500 schools need science teachers and in some cases more than one teacher will be required."

From the Wing Colleges an output of from 250 to 350 science teachers might be expected in each year. Even if the other Colleges could double this figure, the present developments in the Training Colleges, welcome as they are, need to be taken much further if this most unsatisfactory position in the schools is materially to be improved.

REFERENCES

LESTER, F., 1959, '*Contemporary Physics*', 1, 2, p. 142.
Ministry of Education Pamphlet No. 34, " The Training of Teachers " (1957).
Ministry of Education Pamphlet No. 38, " Science in Secondary Schools " (1960).
SCIENCE MASTERS' ASSOCIATION, *Science Teaching in Secondary Modern Schools* (1959).

The Author:
Edward John Wenham is Senior Lecturer in Physics at the City of Worcester Training College. He is a graduate of King's College, London University, and was formerly Senior Physics Master at Sevenoaks School.

Thermoelectric Power Generation

by P. J. BATEMAN
Royal Aircraft Establishment, Farnborough

SUMMARY

Using existing semiconductor materials, it should be possible to produce thermoelectric generators with efficiencies of about 14%. Despite future improvement in material properties it appears unlikely that an efficiency greater than 40% of that of the Carnot cycle can be obtained. This renders thermoelectric devices unsuitable for the large scale production of electric power, and limits their usefulness to certain specialized applications where a suitable heat source and sink are available. When this is the case a generator which would have a power-to-weight ratio comparable with that of conventional small d.c. power sources seems feasible.

1. THE NATURE OF THE THERMOELECTRIC EFFECT

If two dissimilar materials are connected to each other at two points to form an electric circuit, and the two junctions are maintained at different temperatures then, depending on the thermoelectric properties of the two materials, a current will flow in the electric circuit. If a resistive load is now included in the circuit, electrical energy will be dissipated in the load. The practical realization of such a generator will now be considered.

Any material may have associated with it a thermoelectric power α, which is defined as the rate of change of e.m.f. with temperature difference between the junctions when measured against a standard material which forms the other arm of the thermocouple.

The thermoelectric power of a junction is the algebraic difference of the powers α_1 and α_2 of the materials of the constituent arms. Since thermoelectric power is a function of temperature, the thermoelectric e.m.f. of the couple (the Seebeck e.m.f.) is then

$$\int_{T_0}^{T} (\alpha_1 - \alpha_2) \mathrm{d}T$$

where T and T_0 are the temperatures of the hot and cold junctions respectively. It can be shown from thermodynamic considerations, on the basis of the conservation of energy, that this Seebeck e.m.f. can be replaced by a source of e.m.f. π (the Peltier coefficient) at each of the junctions, and a distribution along the conductors of e.m.f. equal to $\tau(\mathrm{d}T/\mathrm{d}x)$. Here, τ is known as the Thomson coefficient, and the term involving $\tau(\mathrm{d}T/\mathrm{d}x)$ is usually small compared with the total Seebeck e.m.f. over the range of temperature differences for which thermocouples may most usefully be employed. The Peltier coefficient π may be defined as the energy dissipated at a junction per unit quantity of electricity flowing across it. If the current passes in the circuit in the direction encouraged by the Seebeck e.m.f. then the Peltier effect will cause thermal energy to be released at the cold junction and absorbed at the hot junction. The effect is reversible, heat being abstracted from the cold junction on passing current in the opposite direction, and this opens up possibilities for refrigeration.

The physical interpretation of thermoelectric phenomena is treated by Stratton (1957) on the basis of the general theory of conduction of an electron gas obeying Fermi–Dirac statistics. The thermoelectric power of a material is defined as the rate of change of Fermi level with temperature. The Fermi level is taken as the energy at which the Fermi function or occupation probability is one half, and it is here measured with respect to some arbitrary and constant zero of energy. It should be pointed out that the distribution of local sources of e.m.f. suggested in the preceding paragraph is purely conventional, and that, whilst π measures the rate at which heat is evolved at a junction, it does not represent a localized e.m.f. at the junction, the Seebeck e.m.f. and the Fermi level really being continuous around the thermocouple circuit.

2. DESIGN PARAMETERS FOR A THERMOELECTRIC GENERATOR

The efficiency η of a thermoelectric generator is the percentage of the rate of consumption of thermal energy Q_a which is delivered to the load as electrical power W. Thus, for a generator of high efficiency, materials are required which have a high value of thermoelectric power α and which have low values of thermal conductivity k, so that for a fixed temperature difference between the hot and cold junctions a minimum of heat will be lost irreversibly by conduction down the arms. For a given load the maximum efficiency is achieved with a generator having minimum internal impedance; this demands thermoelectric materials of low specific resistivity ρ.

It will be shown that the parameter $Z = \alpha^2/k\rho$ deg^{-1} is a convenient figure of merit with which to describe thermoelectric materials. (Some British workers use the square root of this value.) The most suitable metals with $\alpha \simeq 40\ \mu\text{v}\ ^\circ\text{c}^{-1}$ have a value of Z around $0{\cdot}3 \times 10^{-3}\ ^\circ\text{c}^{-1}$ which leads to efficiencies of only 1% for generators using these materials. Thermoelectric generators really only become a practical proposition when efficiencies approaching 10% can be achieved and, with existing semiconductor materials, which have values of Z greater than $1 \times 10^{-3}\ ^\circ\text{c}^{-1}$, this should be possible.

Considering a single thermocouple of internal resistance r connected to a load of resistance R, we have that the efficiency η is given by

$$\eta = \frac{W}{Q_a} = \frac{W}{Q_1 + Q_h - \frac{1}{2}I^2 r} \tag{1}$$

where Q_1 is the rate at which Peltier heat is absorbed at the hot junction at temperature T_1 due to the current I passing through the circuit. This is given by $Q_1 = \pi_{12} I = (\alpha_1 - \alpha_2) I T_1$ assuming here that α_1 and α_2, the thermoelectric powers of the material of the respective arms, are independent of temperature and that π_{12} is the Peltier coefficient at temperature T_1. Q_h is the rate at which heat energy is conducted down the arms from the hot junction to the cold junction at temperature T_0. Thus

$$Q_h = (k_1 S_1 + k_2 S_2)\left(\frac{T_1 - T_0}{L}\right)$$

where k_1, k_2 and S_1, S_2 are the thermal conductivities and cross-sectional areas of the two arms respectively, and L is their length. The term $\frac{1}{2}{}^2Ir$ in eqn. (1) represents half the Joule heat which is dissipated within the thermocouple, and which, as an approximation, may be considered to appear at the hot junction. Now,

$$I = \frac{(\alpha_1 - \alpha_2)(T_1 - T_0)}{R + r}$$

and it follows that

$$W = R\left[\frac{(\alpha_1 - \alpha_2)(T_1 - T_0)}{R + r}\right]^2$$

Substituting these quantities in eqn. (1) and putting $m = R/r$ we get

$$\eta = \frac{T_1 - T_0}{T_1} \cdot \frac{m}{m + 1 + \dfrac{Q_h r}{(T_1 - T_0)T_1} \cdot \dfrac{(m+1)^2}{(\alpha_1 - \alpha_2)^2} - \dfrac{T_1 - T_0}{2T_1}} \qquad (2)$$

For an arbitrary value of the ratio R/r, the maximum efficiency is given when $Q_h r$ is a minimum, and this occurs when

$$\frac{\rho_1 k_2}{\rho_2 k_1} = \left(\frac{S_1}{S_2}\right)^2,$$

where ρ_1 and ρ_2 are the specific resistivities of the respective arms. It follows that

$$Q_h r = (T_1 - T_0)[\sqrt{(k_1\rho_1)} + \sqrt{(k_2\rho_2)}]^2$$

and hence eqn. (2) will be a maximum when the quantity

$$\frac{(\alpha_1 - \alpha_2)^2}{[\sqrt{(k_1\rho_1)} + \sqrt{(k_2\rho_2)}]^2}$$

has its greatest value. This is known as the overall figure of merit of the thermocouple, z, and has dimensions deg^{-1}.

For a given value of the internal resistance r, the maximum delivery of power occurs when $R = r$; but the maximum efficiency occurs for $R > r$, in fact, for a value of R/r given by

$$\left(\frac{R}{r}\right)_{\text{opt}} = \sqrt{[1 + \tfrac{1}{2}z(T_1 + T_0)]} \qquad (3)$$

The gain in efficiency obtained by optimizing the ratio R/r rather than by taking $R = r$ may be up to 1% of the total energy dissipated. The optimum value of the efficiency η_{opt} corresponding to the condition represented by eqn. (3) is given by

$$\eta_{\text{opt}} = \frac{T_1 - T_0}{T_1} \cdot \frac{M - 1}{M + (T_0/T_1)} \qquad (4)$$

where $M = (R/r)_{\text{opt}}$. Thus η_{opt} of a generator is dependent simply upon the properties of the material of the thermocouple as defined by z, and upon the temperatures at which it is worked; it is independent of the number of thermocouples used in the generator as shown later.

The first factor in eqn. (4) corresponds to the thermodynamic efficiency of a reversible engine, and the second represents the reduction of this efficiency due to losses by thermal conduction and Joule heating. It should be noted that both factors are increased by raising the temperature T_1. In order to achieve the maximum possible efficiency then, only one condition has to be satisfied, namely, that the materials chosen should have the highest possible value of z compatible with the effective temperature T_1 of the heat source; that is, $Z(T_1+T_0)/2$ should be a maximum. A plot of optimum efficiency η_{opt} against T_1 for various values of z with $T_0=300°\mathrm{K}$ is given in fig. 1.

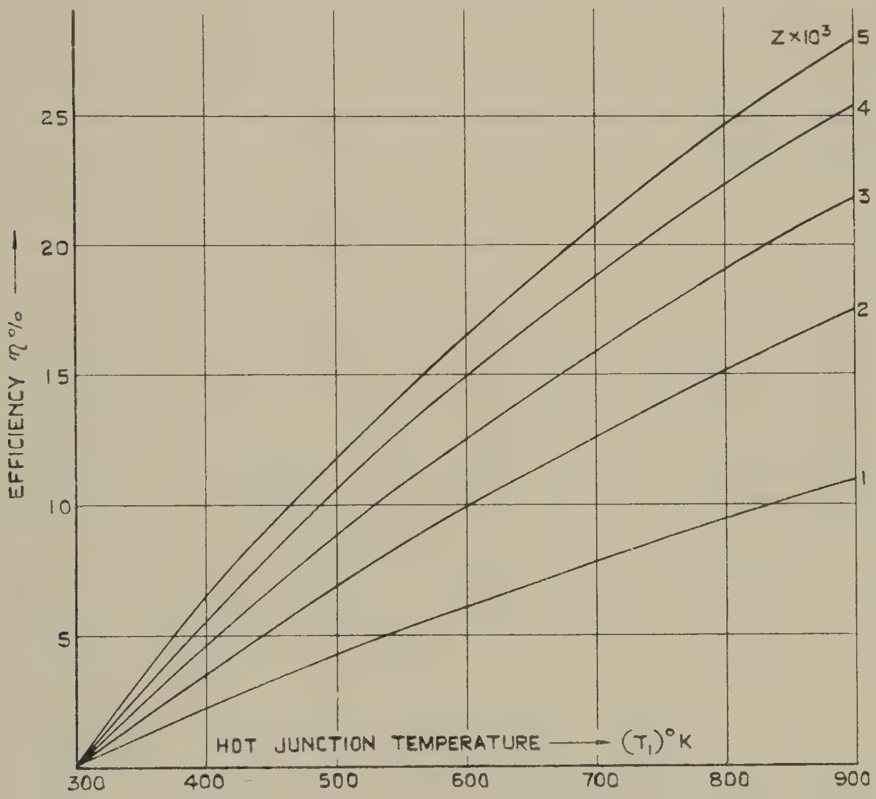


Fig. 1. The effect of hot junction temperature and figure of merit on the efficiency of thermocouples.

Referring to eqn. (1), it is usually found for small currents that the rate of heat conduction Q_{h} is the dominant loss term in the denominator, thus

$$\eta \simeq \frac{W}{Q_{\mathrm{h}}} = \frac{WL}{N(k_1S_1+k_2S_2)(T_1-T_0)},$$

where N is the number of thermocouples comprising the generator. It should be noted that the electrical power W which is dissipated in the load is also proportional to N, so that the efficiency of the generator is independent of the number of thermocouples of which it is composed. Suppose that $k_2=k_1=k$, that $\rho_1=\rho_2$, and that $S_1=S_2=S$. The power output from each thermocouple is then given by

$$\frac{W}{N} = \frac{2kS\eta(T_1-T_0)}{L}.$$

Thus the electrical power generated by each thermocouple depends not upon S and L separately, but upon their ratio. For a given power required this means that the length of the thermoelements is determined by the area available for heating the hot junctions. Also, the power obtainable per unit volume increases inversely as the square of the length L, and is limited only by the restrictions that the electrical resistance of the junctions must be small compared with the resistance of the arms of the thermoelements, and that the temperature drops between the heat source and the hot junctions of the thermoelements, and the heat sink and cold junctions, must be small compared with the temperature drop between hot and cold junctions of the thermoelements. There may also be mechanical limitations on the length L imposed by the brittleness of the material, and the allowable distortion of the generator due to thermal expansion, since the shorter the arms of the thermoelements, for a given temperature difference between hot and cold junctions, the greater the curvature of the slab of elements forming the generator.

3. THE CHOICE OF THERMOELECTRIC MATERIALS

The properties which affect the figure of merit $\alpha^2/k\rho$ are all functions of the concentration of free charge carriers n, that is, electrons or, in the case of some semiconductors, ' positive holes ', and the qualitative variation of these properties with n is shown in fig. 2. The electrical conductivity, $\sigma = 1/\rho$, is roughly proportional to n, whilst the thermoelectric power α tends to zero when n tends to infinity, and α tends to infinity when n goes to zero. But when n is reduced to zero σ must necessarily be very small, and hence the figure of merit may not be increased. The condition is required therefore which will make $\alpha^2\sigma$ a maximum. Figure 2a shows that metals and insulators are inherently poor thermoelectric materials. Figure 2b shows the thermal conductivity k of a substance to be made up of two parts: that due to the free carriers k_{el} and proportional to n, and that due to lattice vibrations k_{ph} (phonons) which is independent of n. The minimum value of k is therefore given by k_{ph}, which is obtained when n is zero. But a small value of n has to be accepted since σ must be as large as possible. Now $\sigma = nu\mathbf{e}$, where u is the carrier mobility and $\mathbf{e}$ is the carrier charge, so carriers of high mobility are required.

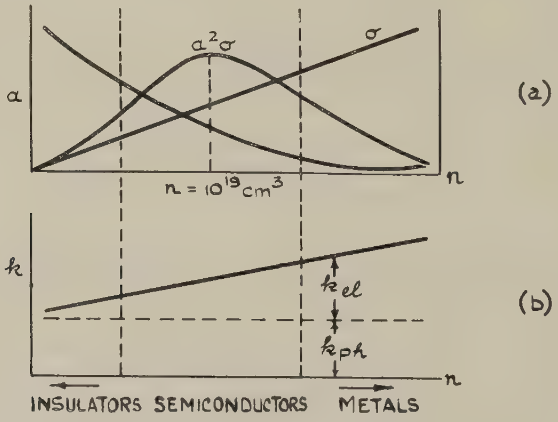


Fig. 2. Variation of material properties with concentration of free charge carriers.

By making the assumption in the electron theory that the maximum value of the figure of merit Z occurs for carrier concentrations which make classical statistics applicable, the Fermi distribution can be replaced by a Maxwell distribution. Also, by assuming that $k_{\mathrm{el}} \ll k_{\mathrm{ph}}$ and that the mobility u is independent of the carrier concentration n, it can be shown that the optimum value of the thermoelectric power is given simply by $\alpha = 2\mathbf{k}/\mathbf{e} = 172\,\mu\mathrm{v}\ {}^{\circ}\mathrm{C}^{-1}$ ($\mathbf{k}$ is the Boltzmann constant), which is the same for all materials under all conditions. The corresponding value of n is of the order of $10^{19}\ \mathrm{cm}^{-3}$, which is a thousand times smaller than for metals and is in the realm of semiconductors. The choice of thermoelectric materials depends then on finding substances which have the highest ratio of u/k_{ph}, and which have a carrier concentration n which gives rise to a value of α around $172\,\mu\mathrm{v}\ {}^{\circ}\mathrm{C}^{-1}$.

The mobility of the charge carriers is limited by their scattering due to thermal vibrations and to lattice defects. The latter can be controlled to some extent by the method of preparation of the materials. The thermal vibrations will be lower at a given temperature, and therefore their effect on mobility minimized, in materials which have a high Debye temperature. Since the scattering is not caused directly by the thermal vibrations of the atoms themselves, but by the disturbances of the periodic potential which they cause, the effect is most marked in ionic compounds and the mobility is therefore greater in materials having binding of a covalent nature. High values of u are found in intermetallic compounds of medium atomic weight. The phonon part of the thermal conductivity can be written in the form $k_{\mathrm{ph}} = \frac{1}{3}cv\bar{\lambda}$, where c is the thermal capacity per unit volume of the material, v is the group velocity of thermal vibrations, namely, the velocity of sound, and $\bar{\lambda}$ is the mean free path for phonons, which is of the order of one or two intermolecular distances in amorphous solids and much greater than the lattice constant in crystals. Low values of k_{ph} are found in compounds and alloys of the heavy elements particularly in the middle groups of the periodic system.

Impurities lower the thermal conductivity of semiconductors by reducing the mean free path of the phonons. When the impurity concentration approaches 50%, $\bar{\lambda}$ falls to a few atomic distances, and in this way the thermal conductivity of a solid solution of two crystals behaves like an amorphous material. Such treatment may however affect the carrier mobility to some extent. Joffe and Stil'bans (1959) describe in some detail the thermoelectric properties of solid solutions.

Summarizing the selection and synthesis of a suitable thermoelectric material: the substance is first ' doped ' with carriers (assuming that there were too few carriers present initially) to give a favourable value of α, that is $\pm 200\,\mu\mathrm{v}\ {}^{\circ}\mathrm{C}^{-1}$, since α^2 in the figure of merit varies more rapidly with n than do σ and k. This treatment may also increase the ratio σ/k, which may be further improved by a reduction of the lattice thermal conductivity k_{ph} by the addition of an impurity compound which crystallises into the lattice in such a way that phonons are scattered, but electrons, with their longer wavelengths (at least an order of magnitude greater than the lattice constant in a semiconductor at 300°K) are not affected, and their mobility is unaltered.

The materials which might be used for thermoelectric generation are intermetallic compounds of Pb, Bi and Sb with Se or Te. The most advantageous arrangement is for the two branches of each thermocouple to be formed from an n-type and p-type semiconductor material respectively. Much work has been done on bismuth telluride, and an alloy Bi_2Te_3–Sb_2Te_3 has now been produced which gives an overall figure of merit z for a p–n junction at room temperature of $2{\cdot}8 \times 10^{-3}$ °C^{-1}. If a material having this figure of merit could be incorporated into a generator with the junctions operated at 600°K and 300°K the optimum efficiency of the device would be a little over 12%.

4. Variation of performance with temperature

For the case where $k_{ph} \gg k_{el}$ the figure of merit will vary with temperature approximately as $Z \simeq f(u/k_{ph} \,.\, T^{3/2})$. In solid solutions such as Bi_2 Te_3–Bi_2 Se_3 and Bi_2Te_3–Sb_2Te_3, which appear to be the most promising for thermoelements, k_{ph} varies little with temperature, whilst the mobility u varies as T^{-2} for temperatures above 250°K. Thus very roughly Z will vary inversely as the square root of the temperature.

As the temperature rises so the electron contribution to the thermal conductivity in a semiconductor increases according to the Wiedemann–Franz Law $k_{el} = (\mathbf{k}/\mathbf{e})^2 A T \sigma$, where A is a constant of the order of 2 for a non-degenerate semiconductor. When $k_{el} \gg k_{ph}$ we have that

$$zT = \frac{\alpha^2 \sigma T}{k_{el}} = \frac{\alpha^2}{A(\mathbf{k}/\mathbf{e})^2} = \frac{\alpha^2}{1{\cdot}3 \times 10^{-8}}$$

from which it can be seen that zT will have a limiting value which will determine the maximum attainable efficiency. If $k_{el} \gg k_{ph}$ then the optimum value of α will be greater than 172 μV °C^{-1} (which is that derived for the condition $k_{ph} \gg k_{el}$), and the limiting value of zT will be somewhat greater than 3, but in practice, material properties do not seem to indicate that these higher values of zT will be obtainable. At present, bismuth telluride and its alloys have shown a value of zT only a little above unity for temperatures below 150°C with a fall off above this temperature, whilst lead telluride alloys appear satisfactory up to 500°C where zT should be near 0·8.

From eqn. (3) it is worth noting that when $zT = 3$, M is approximately 2 which leads from eqn. (4) to a value of η_{opt} between 30% and 40% depending on the relative values of T_1 and T_0. The highest attainable efficiency will therefore be a little less than half the Carnot cycle value.

The maximum working temperature of semiconductor materials will be limited by effects such as atmospheric oxidation, decomposition of the semiconductor, and diffusion of impurities from the contacts. The preparation of satisfactory contacts to withstand high temperatures is also a difficulty. Bismuth telluride alloys are attacked by the atmosphere above 250°C, they begin to dissociate at temperatures around 400°C and they melt near 580°C. Thus it is difficult to specify exactly the maximum working temperature of these materials, but for higher temperatures, refractories, such as transition metal oxides, are indicated in view of their relative mechanical and chemical stability. These, however, appear to have low figures of merit in the region of 1000°C, and at present it seems that the best efficiency is to be gained by using thermoelements in cascade, thereby working the high figure of merit materials in their optimum temperature range.

5. Practical devices

The advantages of thermoelectric generators are that they can be used to convert into useful electrical power heat energy which would otherwise be wasted; they require no servicing during use; they are noiseless, containing no moving parts, and therefore a long operational life may be expected from them. Against this, their low efficiency and the need for a d.c.–a.c. converter for the transmission and distribution of supplies renders them unsuitable for the generation of large quantities of electricity. However, in certain specialized applications such as battery charging, remote warning beacons and weather stations, thermoelectric devices show considerable promise.

In circumstances where a convenient heat source and sink are available and thermoelectric generation is a practical proposition, weight may be at a premium, and it is of interest to compare a thermoelectric device with a conventional battery on this basis. The weight of a hypothetical generator designed to give an output of 4 amp at 25 v is shown in fig. 3 as a function of temperature difference between the hot and cold junctions, the figure of merit for the thermoelement materials, and the length of the thermoelement arms. The form of construction was taken, for the purposes of the calculation, to consist of sintered thermoelement arms

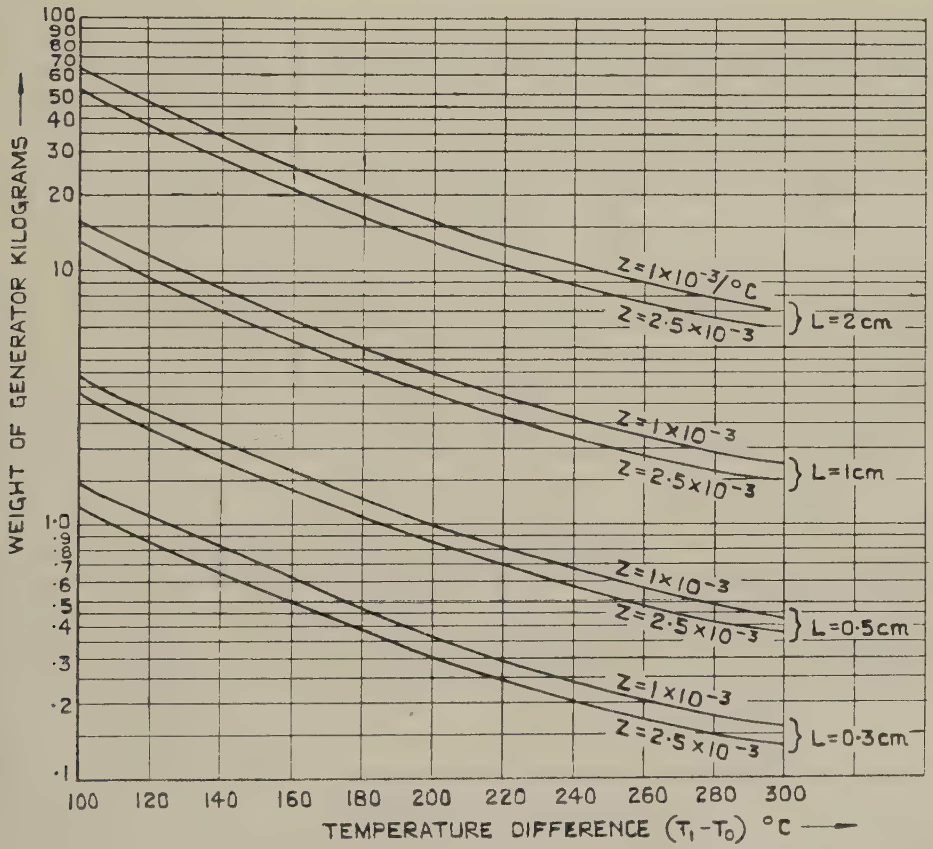


Fig. 3. Design features of a 25 v 100 w thermoelectric generator.

having a ratio $L/S = 4{\cdot}2\ \mathrm{cm}^{-1}$ with an allowance of 50% of the total cross-sectional area for insulation and a 20% increase in the length of the arms to allow for the conducting links and insulation between junctions. A 100 watt generator designed in this way using materials of figure of merit $2{\cdot}5 \times 10^{-3}\ {}^{\circ}\mathrm{C}^{-1}$ and with arms 0·4 cm long supporting a temperature difference of 200°c, would weigh about 600 gm, and would have a surface area of hot junctions of 150 cm^2. By comparison, an existing nickel-cadmium cell, a type suitable for high discharge rates, having a weight of 700 gm and a volume of 200 cm^3, is capable of delivering 120 w at 20 v for only 2½ to 3 min, whilst a 24 v 4 amp-hour lead acid accumulator would weigh 2800 gm.

It should be pointed out that this favourable weight-for-weight comparison is obtained without taking into account the weight of any ducting or cooling fins, for example, which may be necessary for the efficient transfer of heat between the thermoelements and the heat source and sink. Difficulties may be anticipated in the construction and mounting, particularly with the short thermoelement arms needed for minimizing weight, and in the absolute temperature level at which the temperature differential is obtained. Much work has been devoted to these problems in Russia where a generator giving 15–20 w, heated by means of a kerosene burner, has been manufactured for feeding radio transmitters in remote regions, and where a 100 w experimental solar thermoelectric generator is in operation.

The question of adapting the heat source to the generator is of prime importance, and a device described by Barmat, Anderson and Bollmeier (1959), which was designed for powering radio transmitters in satellites and rockets, overcomes this in an ingenious manner. As this device, known as the SNAP III generator, also illustrates the way in which a number of other problems which have arisen in the preceding discussion have been solved, it is appropriate here to describe some of the techniques employed. The heat source is provided by a 3000 curie capsule of polonium 210. The 5·3 Mev alpha particles emitted are absorbed by the hot junctions of the thermoelements which are arranged around the periphery of the capsule as shown in fig. 4. In fact only 26% of the available source area is covered with thermoelectric material which comprised 27 thermocouples. By virtue of its application the heat rejection at the cold junctions is by radiation, and the temperature here fell from 200°c initially to 80°c after a period of 138 days—the half life of the polonium. The corresponding hot junction temperatures were 600°c and 400°c respectively. Initially the power output was 5·3 w at 4·3 v falling to 2·5 w at 3·5 v after 138 days, the overall efficiency being between 5% and 5·5%. The thermoelements 2·5 cm long were of lead telluride doped with bismuth to give n-type material and doped with sodium to give the p-type. The n-type arms were formed in two parts doped differently so that each is operated with maximum effectiveness in the working temperature range. Lead telluride in common with other semiconductor materials is an order of magnitude stronger in compression than in tension, so that the design requires that the stresses be limited to compression. Also, since bonded connections at the hot junctions would be difficult in view of the high operating temperature, both these conditions can be met by using spring pressure applied at the cold junctions, which are themselves soldered, to hold the hot junctions firmly against the wall of the source capsule. The spring pressure also ensures good reliability of the contacts at the hot junctions where sublimation is liable with materials working at their upper temperature limit. In fact, in lead telluride above 480°c plastic flow occurs to give a mechanically bonded contact at the hot junction shoe, thus minimizing the thermal impedance at the interface. The hot junctions are insulated by means of mica plates and an insulating coating separates the cold junctions from the aluminium outer ring A reducing atmosphere inside the sealed unit avoids the risk of oxidation of the thermoelements, and in this way the mechanical bonding of the hot junctions is enhanced by reducing oxides formed on the materials of the interfaces. This effect is reflected in a lowering of the internal impedance of the thermopile which can be detected during the first few hours of operation of the thermopile.

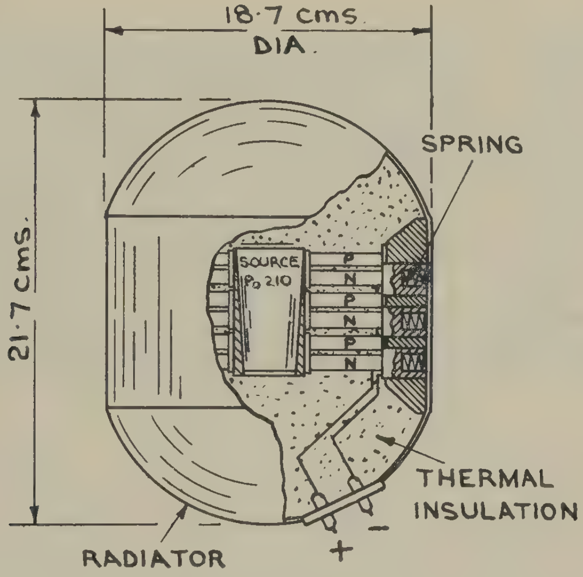


Fig. 4. The SNAP III thermoelectric generator.

During its first half life the SNAP III generator weighing 5 lb gave 9000 w-hours of energy which compares with 250 w-hours from the same weight of commercial dry cells. Using a cerium 144 source with a half life of 290 days, the total energy output per unit weight of fuel would be nearly doubled. Thus in this specialized application, even with an efficiency as low as 5%, a thermoelectric device has been usefully employed for power generation.

References

Barmat, M., Anderson, G. M., and Bollmeier, E. W., 1959, *Nucleonics*, **17**, 166.
Bollmeier, E. W., 1959, *Electrical Engineering*, **78**, 995.
Goldsmid, H. J., 1960, *Applications of Thermoelectricity* (London: Methuen).
Joffe, A. F., 1957, *Semiconductor Thermoelements and Thermoelectric Cooling* (London: Infosearch).
Joffe, A. F., and Stil'bans, L. S., 1959, Reports on Progress in Physics, **22**, 167.
Jaumot, F. E., 1958, Proceedings of the Institution of Radio Engineers, **46**, 538.
Stratton, R., 1957, *British Journal of Applied Physics*, **8**, 315.

The National Chemical Laboratory

OPEN DAY 18TH OCTOBER 1960

The Director, Dr. J. S. Anderson, F.R.S., in his opening address, pointed out that the main research functions of the Laboratory are threefold. Fundamental studies which are important to the advancement of scientific knowledge but inappropriate to universities for various reasons; in this category are the expanding programme of thermodynamic measurements, the hydrometallurgical investigations and the work on new and useful polymers. Basic researches of wide application, as for example on corrosion. And to join with industry in researches of an objective nature. Although a considerable amount of the work leads to empirical results, nevertheless their ultimate interpretation in terms of fundamentals is always borne in mind.

A Steering Committee under the Chairmanship of Professor C. E. H. Bawn, F.R.S., advises on the research programme of the Laboratory, and the work is organized in five main research groups together with a Radioisotopes Applications Unit at Wantage Radiation Laboratories.

1. FIELD-EMISSION MICROSCOPY

This was demonstrated by Dr. J. P. Jones in the Director's Research Unit.

When a sufficiently high voltage is applied to a cold metal, electrons are emitted. Use is made of this phenomenon in the field-emission microscope (fig. 1), in which electrons emitted by a fine metal tip impinge upon a phosphor screen producing a greatly magnified pattern of the tip which depends upon the emissivity of the crystal planes present in the tip. In a region of the surface

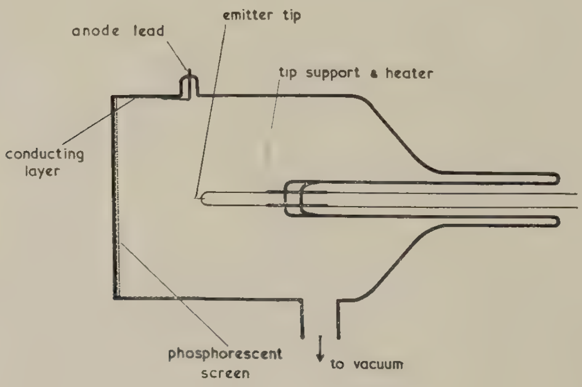


Fig. 1. Principle of the field-emission microscope.

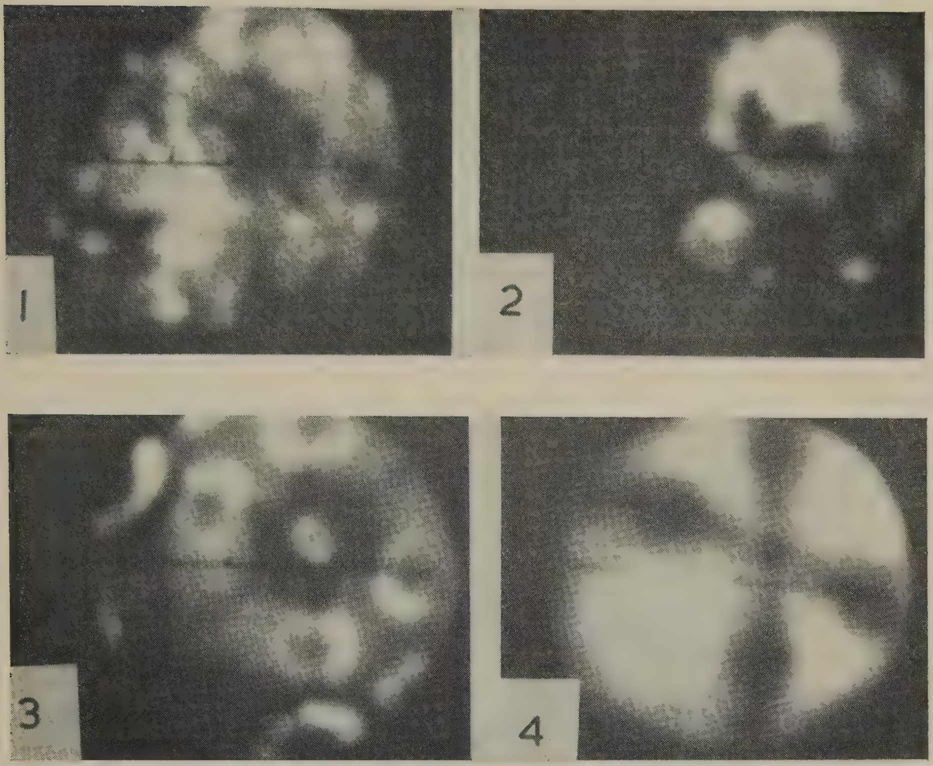


Fig. 2. Field-emission microscope photographs :

1. Nickel condensed as brightly emitting nuclei on the surface of a tungsten tip at room temperature.
2. On warming, nickel atoms migrate over the surface on to preferred crystal planes.
3. Further heating results in migration of metal on to regions which appear as bright rings around the crystal planes of low emissivity.
4. The rings disperse at a still higher temperature and nickel invades most of the surface, leaving uncovered only the planes of low emissivity.

covered by another metal, e.g. nickel, the emissivity is increased, thus the region appears as a bright area on the pattern (fig. 2).

2. THERMODYNAMICS

For the physicist other interesting exhibits were those in the Chemical Thermodynamics Group, whose main activity is the systematic determination of thermodynamic data for organic and inorganic substances. Accurate values for free energy functions, heats of reaction, gas imperfection, covering the widest possible range of substances, are indispensable for pure and applied chemical research, for process development and for chemical engineering. The requisite range of measurements with all the ancillary data, when coupled with the study of methods for the purification and rigorous characterization of the substances used, represents an investment of scientific resources in equipment and man-power, on a scale possible only to a National Laboratory over a period of years. In planning the work it was recognized that there should be co-operation between the few laboratories able to set and maintain the highest standards of precision in measurement, and the Chemical Thermodynamics Group has established contacts with other groups working in the same broad field. The particular choice of programme is that which, after due consultation and with a knowledge of work in progress elsewhere, appeared most likely to cater for the current and foreseeable needs of the British chemical industry.

3. VAPOUR-PRESSURE/TEMPERATURE MEASUREMENTS

The apparatus, nitrogen filled, consists of two identical units connected to the same pressure line (fig. 3). Each unit consists of a boiling vessel, electrically heated (in which the vapour temperature is read on a resistance thermometer) a condenser, and a cold trap immersed in a Dewar flask containing a mixture of acetone and crushed solid carbon dioxide (temperature − 78C°). One boiler contains the substance under examination (e.g. phenols, pyridine bases), and the other contains water whose boiling temperature at a pressure equal to that of the vapour pressure of the substance is read on a resistance thermometer, from which temperature the actual pressure in the apparatus is read from tables.

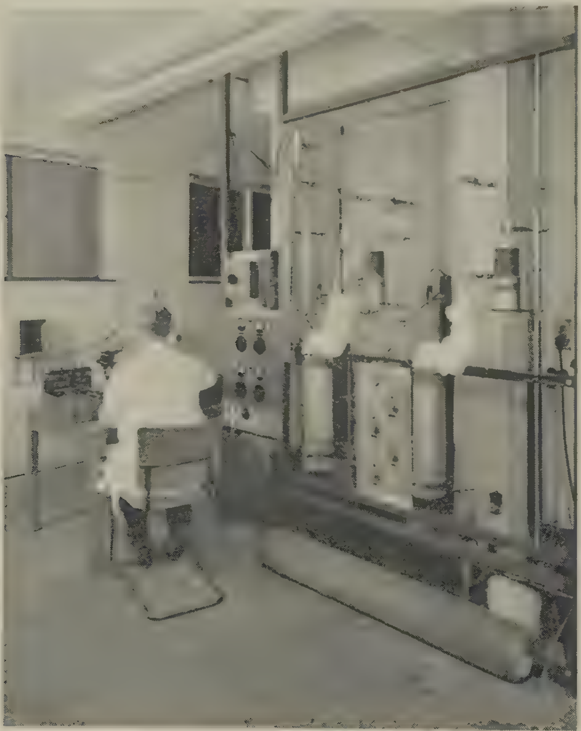


Fig. 3. Apparatus for vapour-pressure determinations.

The demonstrator said that he uses the Antoine empirical equation, $\log p = A - B/(C + t)$ to express the experimental data. It is interesting to note that this equation, unlike the Kirchhoff and other equations, readily lends itself to a change of dependent variable from p to t. (See references 1, 2, 3.)

4. MEASUREMENT OF CRITICAL TEMPERATURES AND PRESSURES

This work was undertaken because a comprehensive review of critical properties showed that the extant data are unreliable for many substances.

The method used was that of observing the disappearance and reappearance of the meniscus between vapour and liquid as the temperature of a sealed tube containing the material passes up and down through the critical temperature. The visitor was able to examine the apparatus and to observe the disappearance and reappearance of the meniscus and some of the interesting phenomena connected with it.

The oven designed for the earlier series of investigations is shown in fig. 4.

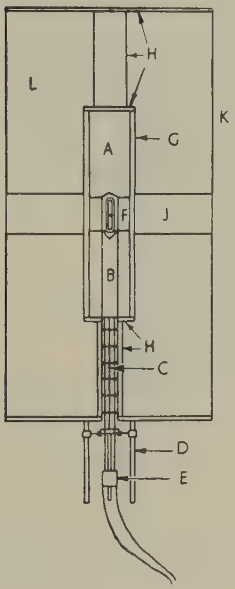


Fig. 4.—Oven for determination of critical temperatures.

A, aluminium block 10 in. long × 1·5 in. diam.;
B, tube carrier, ½ in. diam. dural rod;
C, finned rod to support B;
D, support for B;
E, slide for thermocouples;
F, vertical slots cut in block;
G, copper tube;
H, Uralite;
J, flattened tubes hard soldered to G;
K, brass case 21 in. long × 10 in. diam.;
L, Brelite lagging.

At the beginning of the work nichrome-constantan thermocouples were used for the measurement of temperature, but these have since been replaced by a platinum resistance thermometer used with a Mueller bridge. This equipment is sensitive to 0·001°C, but because of the nature of the phenomenon no greater accuracy than 0·1°C is claimed in the values of the critical temperature.

It is interesting to observe that the values obtained by Young (1885–1910) with mercury thermometers are seldom found to differ from modern values by more than a few tenths of a degree C.

The work has now been extended to cover the measurement of critical pressures and the apparatus is shown in fig. 5.

The oven is seen top-centre and in the foreground is seen the piston gauge for measuring pressure—a development of a type of pressure gauge first used extensively by Amagat. In operation the experimental tube (which communicates with the pressure gauge) is held at the critical temperature in the oven so that the meniscus is in view. Under these conditions a slight increase in pressure (e.g. the addition of 1 g.wt, equivalent to about 1 mm of mercury, to the pressure gauge) leads to adiabatic heating and the disappearance of the

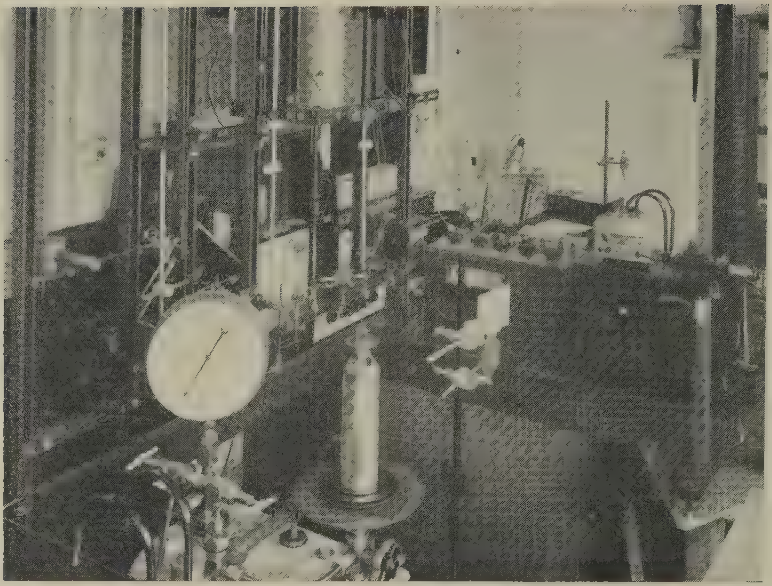


Fig. 5. Apparatus for determination of critical pressure.

meniscus which reappears when the pressure again is reduced. (See references 4, 5.)

5. Measurement of virial coefficients of organic vapours

Taking up an idea first developed by Clausius in the middle of the last century, Onnes, at the beginning of this century, proposed, as the equation of state of a gas,

$$pV = A + Bp + Cp^2 + \text{etc.},$$

where A, B, C, etc., are functions of temperature only and are known as virial coefficients.

In the work carried out with the equipment which was shown it had been established that, for the substances under investigation, Cp^2 was immeasurably small at pressures less than 1 atm. To obtain values of B the essential requirements of the investigation are that the pressures of the vapour, p_1 and p_2, should be determined when a fixed mass of the vapour occupies two fixed volumes, V_1 and V_2, at a chosen temperature T. A full description of the apparatus and the results have been published.

6. Calorimetry

(*a*) *Measurement of heat capacity and latent heat of organic vapours.*

This was a combined determination; the liquid was evaporated by electric heating and then the vapour was electrically heated for the heat capacity determination.

(*b*) *Bomb-calorimetry of organic compounds.*

The static bomb of recognizable design, inspected by the visitor, was interesting, but much more so was the rotatable bomb devised for the determination of the heats of combustion of certain groups of substances, amongst others the halogen compounds (fig. 6). Nearby, in course of construction, there were

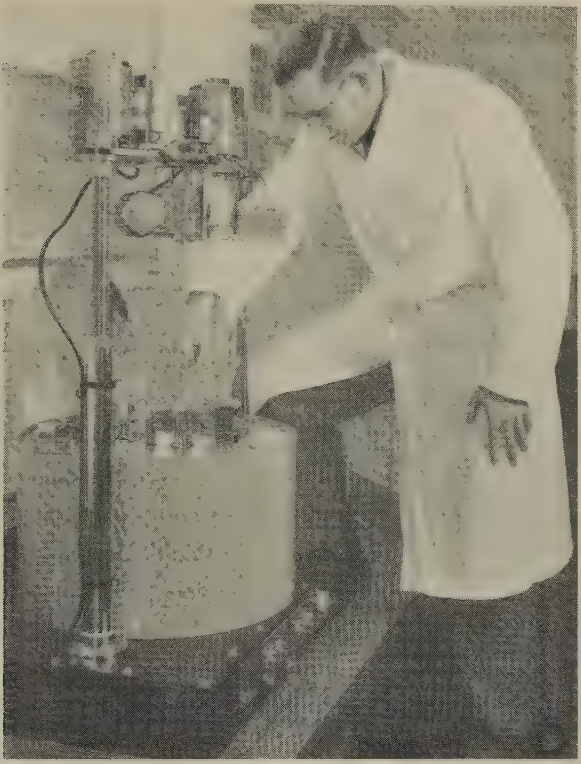


Fig. 6.
Rotating bomb calorimeter.

two helium cryostats for the measurement of the heat capacities of liquid and solid substances down to 10°K or lower.

Other exhibits which the writer had marked at pre-selection in his catalogue as demanding attention, but which the swift passage of time prevented him from reaching were—

Gamma-neutron equipment for beryllium assays.

Separation of the rare earths by ion-exchange chromatography; flow sheets and examples of typical separations.

Electrochemical and radio-active tracer studies of the mechanism of corrosion inhibitors.

Electron diffraction equipment; examination of thin surface films on metals.

V. T. SAUNDERS.

REFERENCES

1. HERINGTON and MARTIN, 1953, *Trans. Faraday Soc.*, **49,** 154.
2. BIDDISCOMBE and MARTIN, 1958, *Trans. Faraday Soc.*, **54,** 1316.
3. COULSON, COX, HERINGTON and MARTIN, 1959, *J. Chem. Soc.*, 1934.
4. AMBROSE and GRANT, 1957, *Trans. Faraday Soc.*, **53,** 414.
5. AMBROSE, COX and TOWNSEND, 1960, *Trans. Faraday Soc.*, **56,** 454.
6. ANDON, COX, HERINGTON and MARTIN, 1957, *Trans. Faraday Soc.*, **53,** 416.
7. COX and ANDON, 1958, *Trans. Faraday Soc.*, **54,** 431.
8. COX, 1960, *Trans. Faraday Soc.*, **56,** 451.

Mathematics for Physicists and Engineers

A Seminar organized by OEEC, February 1961

A number of the papers presented at the OEEC Seminar on *Mathematical Knowledge Required by the Physicist and Engineer*, which was held in Paris from 14 to 17 February 1961, dealt with syllabus content, and with the spirit in which the subject should be taught, reinforcing the arguments and suggestions made in the Institute of Physics pamphlet on the subject which was reissued in 1959. But three of the most stimulating contributions reflected the ' explosion ' (a word used more than once) that has shaken physics and mathematics and psychology into facing the problems of living with high-speed computers.

Dr. H. O. Pollak, of Bell Telephone Laboratories, pointed out that the physicist nowadays tries to master the basic concepts of his subject thoroughly (rather than to learn the details of particular devices), in the hope that he may be able to follow future developments as they emerge. The same applies in mathematics which now needs a fuller understanding of the abstractions with which algebra deals, and a more rigorous approach to fundamentals; the higher we may have to build, the more secure must the foundations be. He mentioned some of the matters discussed, though not officially recommended as yet, by the panel of the Mathematical Association of America that has been examining undergraduate programmes ; these include a course in linear algebra very early in the student's career ; exposure to one rigorous mathematics course such as elementary real variables ; and for the better student a study of the techniques of optimization—linear programming, dynamic programming, game theory, and the calculus of variations.

Mrs. H. J. L. Abercrombie, of Guy's Hospital, spoke on the psychological factors in teaching which make for the effective grasp of concepts. She dealt with the problem of transmitting an established body of knowledge in such a way that the student can use it flexibly in a variety of situations, some of them really new. One wonders whether the average physicist could tell you what the concepts of mathematics are, anyway ; number, and the operations that we can do with numbers ? Newton, in the *Principia*, took the line that mathematics exists to do sums with, and that the nature of the quantities that are being counted is outside its scope. Mrs. Abercrombie's paper suggests that with a suitable emotional approach, people can be guided to make creative use of its processes instead of just ' learning ' mathematics.

But what *are* the processes ? Why is mathematics ? The whole activity of mathematical thought has been examined in many quarters of recent years. One feels that, basically, the latest reappraisal is probably an expression of the need to keep at least one or two steps ahead of the computers. In France this is the *bourbakiste* movement, started by a group writing under the pseudonym ' Bourbaki ' ; they have explored the foundations of mathematics with a new and unrelenting rigour. (C. D. S. Bourbaki, 1816–97, was one of the most distinguished French generals of his time ; his name presumably symbolizes militant zeal. Though it may indicate rigour; or even explosion. Professor M. Jacob (Lyons) discussed the extent to which the new methods and reasoning of this modern mathematics could be introduced to schoolboys, suggesting that probability and the idea of sets might be done fairly early, and that much geometry, conics, inversion, and so on could be omitted from the programme. He mentioned some French textbooks subscribing to these principles, and also ways of relating mathematics to topical and everyday problems. The excellent book by F. Land, *The Language of Mathematics* (John Murray, 1960, 21s), which will be fully reviewed shortly, is probably the first British example of this approach.

Finally, it seems that much of this ground was covered one way or another in the discussions of the Liverpool University Conference on *Mathematics, Education, and Industry*, held in April 1959 (the report, published by *Technology*, Printing House Square, E.C.4, costs 7*s*. 6*d*.). Without a lot of rhetoric, Mr. E. D. Hodge and Dr. T. J. Willmore discussed the nature of mathematical thinking and its relation to the teaching of mathematics; and Mr. B. H. P. Rivett outlined the mathematics involved in operational research, much as Dr. Pollak laid down in his seminar contribution.

Contemporary Reading

The World of Physics, edited by Arthur Beiser with a foreword by E. U. Condon (McGraw Hill, 21*s*. 6*d*.) is a planned course of background reading for the science student. Professor Beiser has really compiled a historical account of the development of the main concepts of modern physics by selecting chapters from the works of different authors (and in one or two cases using complete essays or papers), prefacing each with a short note on the author and his topic. First Galileo's account of his work on falling bodies, which emphasizes that modern science really began with the concept of acceleration ; next, a chapter from Bertrand Russell's *History of Western Philosophy*; the conservation of energy as expounded in Einstein and Infeld's *The Evolution of Physics*; entropy from G. Gamow's *Matter, Earth, and Sky ;* K. Mendelssohn's article from *Science* on Superfluids; Clerk Maxwell's Royal Institution lecture on Action at a Distance, which pays full tribute to Faraday; Lincoln Barnett on Special Relativity; Marie Curie on The Discovery of Radium; M. S. Livingston's memorial lecture to Lawrence, The History of the Cyclotron, recently published in *Physics Today* ; the British Information Service's 1945 account of The Birth of Atomic Energy, specially selected in the belief that the Smyth Report failed to stress the strength of British participation ; E. U. Condon on Quantum Physics ; Abdus Salam on Elementary Particles ; S. A. Korff on Cosmic Radiation ; R. R. Post's famous *Scientific American* article on Fusion Power ; and a short note from the Rand Corporation on Space Research. Condon in his foreword mentions the problem that physics teachers face today—the need to teach basic physics much as we were taught (which, he says, is efficient enough for pupils already motivated to learn it) and at the same time convey the exhilaration of modern ideas and developments. I think that he would accept the traditional formal instruction, and supplement it with reading of the kind that this volume provides. Which brings me back once again to the old sore point, the bogus thesis that you will get better educated scientists by making them do less science ; when, of course, the only way is to make them read more widely about the history, philosophy, and foundations of science. This book is excellent, and should be in every library.

The Advance Science Master's Handbook, by Ivan Muter, who is Senior Lecturer in Electrical Engineering at Rotherham College of Technology (Advance Components Ltd., 12*s*. 6*d*.) is a valuable compilation that should be helpful to the teacher of S-level classes, and useful also at lower levels for the experiments on wave-motion and acoustics. I must be critical of its *aims*, for the introduction states clearly that it hopes to help the science master in his task of producing future electronics engineers and technicians. I hope that he will never be set such a limited task, and that it therefore won't ; but related to what actually goes on in school, it will make some useful additions to the repertoire of experiments. Of the 47 listed, more than half could very profitably be done either as demonstrations or in class. The remainder are either too technical or too difficult. Most of the experiments require a signal generator, valve voltmeter, or vibrator unit, products in which this firm specializes for educational purposes ; and school laboratories, as we emerge from the shoe-string and sealing-wax era, could reasonably expect to have major equipment of this kind. The first half of the book deals generally with theory and circuitry at a rather high level ; it says a lot about transistors, on which there are no experiments offered ; and gives useful information on things like preferred numbers and colour codes.

From the Royal Institution comes *Michael Faraday ; A List of his Lectures and Published Writings*, by Alan E. Jeffreys, with a foreword by Sir Lawrence Bragg (42*s*.). Mr. Jeffreys, who worked in the Library of the Royal Institution, is Senior Assistant at Birmingham University Library. Some 489 items, in chronological order, give a complete bibliography of Faraday's works, and the list gives some idea of his industry and breadth of interests. The book is well produced with a number of excellent photographs.

Eléments de Physique Nucléaire, by Daniel Blanc and Georges Ambrosino (Paris, Masson et Cie, 30 N.F.) is a straightforward presentation of nuclear physics for engineers and 'professeurs non spécialistes', which brings it into my clutches. It is very good in explaining matters like the use of centre-of-mass co-ordinates, how cross-sections are reckoned and measured, and nuclear resonances. Recent ideas on the structure of nucleons, and work on mesonic atoms, are mentioned. Fission reactors are explained without technical details, and the chapter on fusion has, as the reference indicates, relied to some extent on Post's *Scientific American* article. The book is fully documented with references. The price, 30 N.F., seems rather high for a paper-back ; and the latest edition of Glasstone would, one must admit, be a better buy as a reference book.

As a work of scholarship and exposition, Sir Harrie Massey's *The New Age in Physics* (Elek Books, 42*s.*) is impressive. Its aim is to explain in terms ' as non-technical as possible ' what modern physics is about, and it is really a chronicle of the developments of the last ten years or so, prefaced by an account of the evolution of the main ideas that led to this proliferation. Non-technical means without a lot of wave-functions and operators; in fact it demands as close and careful reading as any worth-while book on physics, and is not for the dilettante or beginner. Anyone who graduated in physics before 1950 will learn a great deal from it ; everyone who is teaching physics can be strongly recommended to read it, and to make it available to his pupils. The chapters on radioastronomy, the upper atmosphere, and artificial satellites and space probes, cover the more spectacular advances admirably. What is unusually good is the treatment of fundamental problems—electrons in solids, the self-energy of the electron and the Lamb Shift, the maser, strange particles and strange atoms, the neutrino, antiparticles and the uncertainty principle.

Controlled fusion power may be some way off yet, but what do we do when it comes? This question is discussed in *The Challenge of Fusion* by Duncan Curry and Bertram R. Newman (Van Nostrand, 41*s.* 6*d.*). The first part of the book deals with Project Sherwood, and can for the purpose of this article be called the last Post. Then the impact of unlimited power on social life in small communities and in the world at large is considered. The moral, that people released from much of the struggle for bread will have to learn how to live, leads to an urgent plea for planning *now* a balanced general education to raise the standard of living in human rather than material terms.

Nuclear Propulsion, edited by M. W. Thring (Butterworth, 50*s.*), can be recommended for the library.Three of the sixteen chapters are technical, and the rest is a clear and easily readable account of nuclear plants in ships, aircraft, and possibly space vehicles, the manipulation of working fluids, new materials, and the accommodation of the crew.

The P.S.S.C. textbook *Physics*, with the laboratory guide, is now published in Britain by Harraps (40*s.* and 10*s.* respectively). Both have been revised in the light of experience in action. Teachers will find these stimulating and rewarding books to read.

The original American edition has already been reviewed (Vol. 1, No. 3, 1960), and the aims of the Course were discussed by Mr. Stephen White of P.S.S.C. (Vol. 2, No. 2, 1960). The British edition has been enthusiastically received, particularly by university physicists who can see that this is the way in which the subject should be taught. Schoolmasters, while approving heartily in principle, will wonder how to inject it into any present or likely grammar-school course. The Science Masters' Association and the Association of Women Science Teachers have just issued suggestions for new school syllabuses (*Physics for Grammar Schools*. Published for the S.M.A. and A.W.S.T. by John Murray, 1961. 2*s.*), which match the P.S.S.C. in policy, aim, and general content. The outcome that one would like to see is a new edition adapted specifically to British needs, arising from a fusion between the material offered in the present edition and the finally accepted shape of the S.M.A. recommendations. Indeed, when one remembers Professor Zacharias' insistence that the textbook is only one of the wide range of teaching aids associated with the Course, such incorporation is to be expected generally as its influence spreads to other countries which already have highly developed science teaching traditions. But there are, I understand, plans to try out the Course in its entirety here ; and the results of this valuable experiment will be awaited with great interest. G.R.N.

BOOK REVIEWS

Classical Electricity and Magnetism. By E. S. SHIRE. (London: Cambridge University Press, 1960.) [Pp. 396.] 45*s.*

This book is at the level normally taught during the first year of an honours physics course. It is obviously based on a course Mr. Shire has been giving at Cambridge. It will be of particular interest to those teachers training pupils for Cambridge scholarships as it gives an insight into the way electricity and magnetism are taught at Cambridge. To quote the preface, the book endeavours to "bridge the gap between an elementary first account of electricity and a formal mathematical treatment that starts by accepting Maxwell's equations as true". Unlike school textbooks on electricity it does not deal with thermoelectricity, electrolysis, atomic and modern physics as these subjects are normally taught in complementary courses. After an introductory chapter, there are three chapters on electrostatics. Then follows a chapter on magnetostatics that in the space of thirty pages covers the forces between magnetic poles, magnetometry, geomagnetism, dia-, para- and ferromagnetism, ferromagnetic domains and the design of permanent magnets. The magnetic effects of currents are introduced in Chapter 6, which includes an excellent account of Ampère's experiments. Following a chapter on electrical conduction there is a chapter on electromagnetic induction. The difference between a motional e.m.f. and the e.m.f. induced in a stationary circuit by a varying magnetic field is not emphasized. Chapter 9 is on circuit theory and includes a discussion of transients, A.C. theory and network theorems. Chapter 10 gives a brief introduction to electromagnetic waves. Chapter 11 is on electrical measurements and gives an excellent account of units, absolute measurements and the maintenance of standards. Chapter 12 is on applied electricity and includes accounts of dynamos, motors, electronics, the motion of charged particles and particle accelerators.

There has been a great deal of controversy recently about the teaching of electricity and this book makes several interesting contributions. Mr. Shire writes "I make no apologies for giving magnetostatics a prominent place. Nor do I shrink from the use of magnetic poles as some would have us do; properly employed, with the appropriate physical restriction in mind—that the total pole on an isolated piece of magnetic material is zero—the concept of magnetic pole is as sound as is that of a point electric charge". Mr. Shire adds "My views on this and other aspects of teaching lie close to those expressed by Heckstall-Smith (1957) and Macfadyen (1958)". Mr. Shire develops electromagnetism from magnetic poles defining H in terms of the force on a magnetic pole. The weakness of this approach is that it is based initially on H not B as the primary field vector. The alternative approach, preferred by the reviewer at University level, is to start by defining B in empty space in terms of the force on a moving charge; but this approach has the disadvantage that it is a little more complicated, initially as it involves the use of the vector product. The reviewer would agree entirely with Shire when he writes "Within an historical framework there still remains some choice of method—Coulomb's experiments or Cavendish's to justify the inverse square law, magnetic shells or the force between currents to relate currents and magnetic fields. I think that this freedom of choice should be realised and appreciated; no one system of development should be asserted dogmatically to be the only correct one. I have myself chosen a particular route (with an occasional hint of an alternative path), but I should claim no more for it than that I believe that it is consistent, keeps in close touch with experiment and makes it possible to discuss difficulties as they arise". The above quotation typifies the spirit in which the book is written.

The other interesting feature of Shire's approach is that no one system of units is used exclusively. Shire writes "a physicist should be able to move with ease from one system to another, choosing whichever is the most convenient for the purpose in hand". Coulomb's law is written as

$$F=\frac{\delta}{4\pi\epsilon_r\epsilon_0}\;\frac{Q_1 Q_2}{r^2},$$

and the values of δ and ϵ_0 for five systems of units are given in an appendix. One wonders if such generalization at this stage leads to more confusion than it seeks to avoid. The book gives an excellent account of the historical development of the subject. The development of the theory is always related to experiments, the accuracy of which is then discussed critically. It is certainly a book that will provide a firm foundation for the development of the subject at University level. The book can also be recommended to teachers and scholarship candidates at school as a reliable reference book and as a source for the history of electricity and magnetism.

Some readers may wonder if it is Cambridge's answer to Bleaney and Bleaney. The answer is 'No' as to a large extent the books complement each other, Bleaney and Bleaney taking over where Shire leaves off. W. G. V. Rosser.

Lectures in Theoretical Physics, Vol II. (Interscience Publishers, 1960.) [Pp. 483.] 68*s.*

It was said by Condon that " a technical volume is not worth reviewing if its author has not considered it important enough to provide it with an index". In the case of an edited volume one might perhaps occasionally break this rule as it seems unfair to visit the sins of the editors on the contributing authors. One wonders, however, why the publishers of this volume—which in many respects is an admirable one—having omitted an index and having elected for off-set print (surely, the time factor cannot have been important as more than 15 months elapsed before publication) decided to price the book as if it were a normal production.

Having voiced these criticisms the present reviewer can state immediately that it will certainly come in very useful as a reference book for many of the topics treated by the lecturers at the 1959 Summer School for Theoretical Physics held at the University of Colorado. In a short review one cannot go into details and it must suffice to list the contributors and their topics ; Salam on Elementary Particle Physics, Sakurai on Elementary Particle Interactions, Downs on Hypernuclei, Rohrlich on the Classical Electron, Hugenholtz on the Many-Body Problem, Jensen on Beta-Decay, Dresden on Abstract Field Theory, and Lamb on Masers. The level is on the whole quite advanced, but the discussion is thorough and usually detailed. D. ter Haar.

General Properties of Matter. By C. J. Smith. Part 1 of A General Degree Physics. 2nd Ed. (Edward Arnold Ltd., 1960.) [Pp. 732.] 63*s.*

Dr. C. J. Smith is well known to most British physicists as a prolific writer of text books at various levels. The present volume is the second edition of the first part of his General Degree Physics. The subjects treated include a mathematical introduction, particle and rigid body dynamics, simple harmonic motion, bending of beams, compressibility, surface tension, viscosity, hydraulics, plasticity, Brownian movement, and vacuum practice. The treatment to the standard of the B.Sc. General Degree is thorough and reliable and the success of the first edition will undoubtedly be continued with the present revised version. Subjects such as non-Newtonian fluids which now form an important topic in general degree courses are treated thoroughly.

One must however apply perhaps more critical standards to established authors than to authors of first books, who normally need encouragement if they show any writing ability. Dr. Smith states in his preface that " the ground covered is just a little more than that usually required for the B.Sc. General Degree ". It is in some of the additional material that the book can be criticised. For example, pages 630–640 deal with plastic deformation of solids and the word *dislocation* which provides the clue to an understanding of the slip processes is not mentioned. The treatment of this topic is out of date and since the material is probably additional to that required for the General Degree, the book would be both shorter and better without it. This omission is the more surprising since some of the definitive experiments on the subject were carried out in the laboratories of which the author is Assistant Director. Similar criticisms can be made in the chapter on vacuum practice. The Sprengel and Geissler pumps, now largely of historical interest, are given fairly detailed treatment whereas the ion pump for obtaining really high vacua, (now a normal laboratory and even industrial instrument) is ignored. It is to be hoped that in any future editions the ' additional ' material will be pruned and re-written in a shorter and more up-to-date form. C. A. Hogarth.

Small-scale Preparation of Fine-grain (Colloidal) Photographic Emulsions. Notes on Applied Science Series, No. 20. (Published for DSIR by HMSO.) 1*s*. 6*d*. (U.S.A. 27 *cents*).

With an increasing interest in microphotography, particularly in the fields of precision scales and graticules, scales of the diffraction-grating type and interference filters, it is important to know more about possible methods of making small quantities of photographic emulsions of the finest possible grain size and of the highest quality.

In this latest addition to the NPL series of ' Notes on Applied Science ', Dr. B. H. Crawford, Head of the Photometry and Colorimetry Section of the Light Division, gives a detailed account of a special method he has developed at the National Physical Laboratory for making an ultra fine-grain emulsion.

From Dualism to Unity in Quantum Physics. By ALFRED LANDÉ. (Cambridge University Press, 1960.) [Pp. xvi + 114.] 18*s*. 6*d*.

This is an interesting and unusual book, which attempts to put quantum physics on a new kind of basis. The author disapproves in vigorous and sometimes aggressive language of the widely accepted view that ". . . the same physical object appears in two seemingly mutually exclusive forms, particle or wave . . ." To him the particle is the real thing, the wave merely a picture of the formula by which we compute the probability of finding the particle in different places. To treat the two on equal terms—he says—is like treating on equal terms the individual sufferers from cholera and the curve that depicts the rise and fall of the epidemic.

So much is philosophy; but the author, a physicist of long and respected standing, does more. From certain eminently plausible assumptions about the continuous and consistent behaviour of probabilities he derives mathematically their ' interference ', the Schroedinger equation, and indeed the whole basis of quantum mechanics including the quantum ' postulate '. That, I feel, is an important achievement.

Let me say again: Landé does not try (like Bohr or de Broglie) to explain causally the statistical behaviour of atoms; on the contrary, he shows very plausibly that a probabilistic basis of physics is indispensable. He deliberately leaves out relativistic quantum theory and only briefly touches on photons (which he does not regard as real particles). It will be interesting to see if his ideas will prevail and can be applied to the whole of quantum theory. Altogether this is a most thought-provoking book, and brilliantly written. O. R. FRISCH

Interscience Monographs and Texts in Physics and Astronomy. Vol. IV.—Fast Neutron Physics, Part I. (Techniques). Edited by J. B. MARION and J. L. FOWLER. (New York and London: Interscience Publishers.) [Pp. xiii + 983.] 218*s*.

This book is another example of the composite volume containing review articles, each complete in itself, and in this case the result is wholly good. Fast neutron physics is a field whose technology is rapidly expanding, and where the emergence of reliable nanosecond pulse techniques is enabling the time-of-flight methods so successful in slow neutron physics to be used in the fast neutron field, bringing with them order-of-magnitude improvements in the accuracy of measurements. The production of an up-to-date reference book is therefore all the more difficult and it is much to the credit of the editors of this one that reference to work appearing as late as 1959 has been included.

The articles include various methods of generating fast neutrons, detection and energy measurements, a review of the difficult subject of absolute flux measurement, and finally, such specialized themes as shielding, collimation and radiation hazards involved. Each subject is dealt with in a scholarly fashion by an acknowledged expert and with great attention to detail (the book contains some 983 pages of small print). An attractive feature is the inclusion of many numerical tables of immediate practical use to anyone planning an experiment or designing a laboratory. In short this is a book which any scientist practising or studying fast neutron physics must consult and which should certainly find a place in the library of every nuclear physics laboratory. It is much to be regretted that its price appears likely to prevent all but a few individual workers from owning a copy. M. J. POOLE.

Physics and Medicine of the Atmosphere and Space. Proceedings of the Second International Symposium on the Physics and Medicine of the Atmosphere and Space held at San Antonio, Texas, November 1959. Edited by O. O. Benson, Jnr., and H. Strughold. (New York: John Wylie & Sons Inc., 1960.) [Pp. 645.] 100*s.*

The material in this book can be roughly divided into three classes, the fundamental physics of the upper atmosphere and space, the engineering problems of rocketry (propulsion, guidance and tracking), and the biological problems of survival and behaviour of man in the new environment. Most of the papers are of a high scientific character. Many contain much information gained from the earlier American satellite launchings. The authors, as might be expected, are almost entirely American, exceptions being one paper each from this country, Switzerland and Western Germany.

Many of the papers make fascinating reading. The actual temperatures transmitted by Explorer III and recorded by stations in different parts of the world, the possibility of applying nuclear fusion to rocket propulsion, the fundamental difficulties involved in designing a photon rocket, the effect of weightlessness on the space traveller, " time dilation and the astronaut ", the problems of rescue from space by a second vehicle, are topics which will have a strong appeal for all science fiction addicts as well as to serious workers in the field.

A number of authoritative surveys are invaluable to the general reader and help him to find his way amongst the many specialist articles. The survey papers include, for example, "A Survey of Possible Propulsion Systems " by Dr. L. R. Shepherd, who is Chairman of the British Interplanetary Society and Deputy Chief Scientist at the Reactor Division at Harwell; " Radiation Hazards of Space Flight " by Dr. J. A. Van Allen of Iowa State University, one of the discoverers of the " radiation belts " round the earth; and " The Timing of Satellite and Space Probe Launchings " by Dr. W. von Braun of the U.S. Army Ballistic Missile Agency.

Considering the excellent standard of production, the price of this book is reasonable.

A. Nightingale.

Advice to a Lecturer. By Michael Faraday. (The Royal Institution, 1960.) [Pp. 12.] 2*s.* 6*d.*

Faraday was the foremost scientific lecturer of his time. Contemporary reports stress the clarity of his thought and the enthusiasm he aroused in his audiences. Yet " his manner was so natural that the thought of any art in his lecturing never occurred to anyone " (Bence Jones).

That he had given much attention to this art many years before he began to lecture at the Royal Institution is apparent from a remarkable series of letters to his young friend Abbott, written in 1813, when Faraday was only 21. These contain the substance of his observations on lecturers and lecturing. The present pamphlet is an anthology taken from these letters, the greater part of which originally appeared (in 1870, not 1869) in *The Life and Letters of Faraday* by Bence Jones. It is published in the hope that Faraday's advice may be of help to would-be lecturers today. All practitioners of the art would do well to heed his precepts, whether their subject be physics or philology. May we hope for a companion anthology of Faraday's advice on note-taking? This too would be pertinent today.

A. E. Jeffreys.

Particles, Field Theory and Force

by R. J. BLIN-STOYLE
Fellow of Wadham College and Senior Research Officer in Theoretical Physics, University of Oxford

SUMMARY

An attempt is made to give a simple account of the basic ideas of quantum field theory. This theory provides a description of the way in which the elementary particles of nature acquire certain of their properties and also accounts for their interactions with one another. Particular emphasis is laid on the properties of the neutron and proton, nuclear forces and the electromagnetic properties of the electron.

Some readers may find certain conceptual ideas presented in this article rather difficult to assimilate, the sections 6(*b*), 8, and 9 being particularly difficult in this respect. The article is so constructed, however, that it can be read in a sensible fashion omitting these sections. This course is recommended to the faint hearted!

1. INTRODUCTION

The way in which pieces of matter interact with each other has been a subject of diverse speculation for thousands of years. Such interactions are usually referred to as forces, and in the days before Galileo it was commonly believed that force was necessary to sustain motion. That a body could remain in motion without the continual application of a force seemed unthinkable. Galileo's proposals that a body continued in a state of *uniform* motion when no forces were applied and that a *change* in the motion resulted from the application of a force were indeed revolutionary. These ideas, of course, are now well established and are incorporated in Newton's three laws of motion:

(1) Every body continues in its state of rest or of uniform motion in a straight line except in so far as it may be compelled to change that state by the action of some external force.

(2) Change of motion is proportional to the force applied and takes place in the direction of the line of action of the force.

(3) To every action there is always an equal and opposite reaction.

Perhaps the first most outstanding application of these laws was to the mechanics of the motion of the earth and other planetary objects around the sun. Here it was proposed (by Newton) that between two bodies of mass m and M, separated by a distance r, there is a force $F(r)$ of magnitude

$$F(r) = \gamma \frac{Mm}{r^2} \tag{1}$$

where γ is the so-called gravitational constant and determines the strength of the force. Using this force and Newton's laws it was found possible to account almost completely for the motions taking place within the solar system.

---

†This article was mainly written whilst the author was Visiting Professor at the University of California, La Jolla, California, U.S.A. The hospitality of the Physics Department is gratefully acknowledged.

A force of this nature has considerable conceptual difficulties associated with it, however, in so far as it is a force acting between bodies *not in direct physical contact*. It is an intuitively acceptable idea that one body touching another can exert a force on it. Not so when there is no obvious contact between them and we have what has been called *action at a distance*. This type of interaction had been known for some time and occurs, for instance, in magnetic and electric phenomena. Many attempts were made to give a straightforward, mechanistic interpretation of such a force, almost always in terms of a hypothetical aether. This remarkable ' fluid ', supposedly pervading all space, was intended to give some sort of direct, physical line of communication between interacting bodies. Immense ingenuity was used in order to invest the aether with the requisite properties necessary to account for the known interactions. In spite of this, all such attempts failed and, as is now well known, with the advent of the Michelson–Morley experiment and relativity theory, it was shown quite unequivocably, that there is no place for an aether in the physical world.

Thus, we are still left with the problem of understanding how two bodies, not in obviously direct, physical contact can nevertheless interact with one another. It is one of the main objects of this article to try and give a picture of the currently accepted ideas on this subject. The formulation of these ideas is referred to in the literature as quantum field theory. Before embarking on this, however, it is useful to give an alternative mathematical representation of the force between two bodies in terms of the potential energy concept.

Consider two point particles such that when they are separated by a distance r there is a force $F(r)$ between them. Now let us calculate the amount of energy, $V(r)$, which has to be expended in bringing one particle from infinity up to a distance r from the other. Clearly it is given by

$$V(r) = -\int_{\infty}^{r} F(r')dr' \tag{2}$$

or conversely

$$F(r) = -\frac{dV(r)}{dr} \tag{3}$$

$V(r)$ is referred to as the potential between the particles and for our purposes is a more useful concept to discuss than that of the force itself, although the latter can be obtained from the former by means of eqn. (3).

In order to discuss the way in which two bodies can interact with each other we must now be more specific about the meaning of the word ' body '. A solid state physicist or a chemist, for example, is concerned in his work with the interactions that operate between atoms or molecules—van der Waals forces, exchange forces, etc. Now atoms and molecules are constructed of nuclei and electrons, the nuclei themselves being complexes of neutrons and protons and these ' chemical ' forces can all be satisfactorily accounted for by regarding them simply as the bulk effect of all the interactions (in this case electrostatic) acting between the constituent particles of the atoms and molecules. Thus the obvious first step to take is to try and give a description of the way in which it is believed that two elementary particles(1) (neutrons, protons, electrons, etc.) interact with one another. This article will therefore be concerned *only* with these basic elementary particle interactions and no further discussion will be given of the way in which these interactions manifest themselves when dealing with complexes (e.g. atoms or molecules) of many elementary particles.

It is now necessary to discuss the decay properties of elementary particles which are relevant to the force problem.

2. DECAY OF ELEMENTARY PARTICLES

In an earlier article(2) concerned with the structure of the atomic nucleus it was pointed out that the phenomenon of β-decay, in which a nucleus of charge Ze and mass number A changes, for example, into a nucleus of charge $(Z+1)e$ and mass number A with emission of an electron and a neutrino, had to be interpreted as the spontaneous decay of a neutron into a proton, electron and neutrino. Symbolically

$$n \rightarrow p+e+\nu \tag{4}$$

This is only one of many examples of spontaneous decay of elementary particles. Thus the negatively charged π-meson ($\pi-$) decays into a μ-meson ($\mu-$) and a neutrino; the μ^- decays into an electron (e^-) and two neutrinos. Similarly for the positively charged π-meson (π^+). Again, such processes can be written symbolically as

$$\pi^\pm \rightarrow \mu^\pm+\nu \tag{5}$$

$$\mu^\pm \rightarrow e^\pm+\nu+\nu. \tag{6}$$

Processes such as these are completely different in nature from the emission of α-particles from a nucleus. In the latter case the α-particle is not an elementary particle in the usually accepted sense, but is a complex structure of four nucleons (two neutrons and two protons) and the decay is simply a breaking off of these four nucleons from the parent nucleus to form an α-particle and another nucleus whose mass number is reduced by 4. In a process such as β-decay, however, a neutron is annihilated and a proton, electron and neutrino, *which were not present in any real, physical sense before*, are created. The process is one of spontaneous *annihilation* and *creation* of particles.

So far we have only considered processes of this kind in which one particle is annihilated and different particles are created. However, we can also have the situation in which a particle in one energy state is annihilated and the same type of particle in a different energy state is created together with some other particle(s). The simplest example of this kind is the emission of light from an excited atom. This type of process was discussed briefly in a previous article(2) and is one in which an electron jumps from one energy level A (say) to a level B (say) and emits a quantum of electromagnetic radiation (a photon, γ). In the language of creation and annihilation we can say that an electron in state A is annihilated and an electron in state B is created together with a photon. Symbolically,

$$e_{\mathrm{A}} \rightarrow e_{\mathrm{B}}+\gamma. \tag{7}$$

This sort of process could not happen if the states A and B were the same since there would be no energy available for the photon whereas it is well established experimentally that in all real physical processes, energy is conserved. Energy has to be provided for the process to take place and in the case under discussion it comes from the excitation of the atom.

Another process of this type is that in which a nucleon (neutron or proton) changes its state and emits a π-meson (pion). The following is an example of such a process

$$n+p \rightarrow n+n+\pi^{+}. \tag{8}$$

Here a neutron of high energy strikes a proton and provides sufficient energy for the proton to be annihilated and another neutron and a π^{+} to be created. The bombarding neutron has to have high energy in order to create the π^{+}. A π^{+} has mass 274 m (where m is the mass of an electron) and using Einstein's relation between mass and energy $E=mc^2$ it follows that the bombarding neutron must have an energy of the order 600 MeV† in order for the process to go at all.

It is to be noted that in the decay processes (4), (5) and (6), the sum of the masses of the product particles is always less than that of the parent particles and so there is no problem of providing sufficient energy for the decay to take place—the decay is spontaneous. Not so with processes (7) and (8). The real spontaneous emission of a photon by an isolated electron or of a pion by an isolated nucleon does not take place. Energy has to be provided in some way for the physical emission of photons or pions to take place.

One other point has to be noted, namely, that processes of the foregoing type can also go in the reverse direction. For example, a proton can absorb a π^{-} to form a neutron

$$p+\pi^{-}\rightarrow n \tag{9}$$

or an electron can absorb a photon and go from a state B (say) to a state A

$$e_{\mathrm{B}}+\gamma\rightarrow e_{\mathrm{A}} \tag{10}$$

Furthermore, in a process of the type (9) a fermion (in this case n or p) can be taken from one side of the ‘ equation ’ and replaced on the other side by its corresponding anti-particle. We can have, for example,

$$\pi^{+}\rightarrow p+\bar{n} \tag{11}$$

in which a pion (π^{+}) is annihilated and a proton (p) and an anti-neutron ($\bar{n}$) are created. Similarly,

$$\begin{aligned} p+\bar{n}&\rightarrow \pi^{+} \\ \gamma&\rightarrow e+\bar{e} \\ e+\bar{e}&\rightarrow \gamma \end{aligned} \tag{12}$$

where the bar ($^{-}$) over any particle symbol indicates the anti-particle. The concept of anti-particle has become a familiar one in recent years, mainly through the experimental verification of the existence of the anti-proton and anti-neutron. An anti-particle has the same spin and mass as the particle, but its electric charge and magnetic moment are opposite in sign. Thus the anti-proton is negatively charged; the anti-electron (conventionally called the positron) is positively charged and the anti-neutron has zero charge.

---

†1 MeV (million electron volts) is the energy an electron acquires after being accelerated through a potential change of 1 million volts.

Summarizing then, provided energy can be made available, the following real physical processes can take place

$$\begin{array}{ll} e \rightleftharpoons e+\gamma & \gamma \rightleftharpoons e+\bar{e} \\ p \rightleftharpoons n+\pi^{+} & \pi^{+} \rightleftharpoons p+\bar{n} \\ p \rightleftharpoons p+\pi^{0} & \pi^{0} \rightleftharpoons p+\bar{p} \\ n \rightleftharpoons n+\pi^{0} & \pi^{0} \rightleftharpoons n+\bar{n} \\ n \rightleftharpoons p+\pi^{-} & \pi^{-} \rightleftharpoons n+\bar{p} \\ p \rightleftharpoons p+\gamma & \gamma \rightleftharpoons p+\bar{p} \end{array} \qquad (13)$$

where a number of other analogous processes has also been included in the list. The list is still by no means exhaustive but suffices for illustration—some other processes will be mentioned later.

3. COUPLING CONSTANT

In the previous section various processes have been described in which particles are created and annihilated. We now have to discuss the probability of such processes taking place given that they are allowed energetically. The probability is governed by various factors; the kinematics of the process, the number and masses of the particles involved, the energy available and the nature and strength of the interaction responsible for the process. The first four factors are determined at once by the particular process under consideration and have no great physical interest; the fundamental physics stems from the interaction. Even so, we shall not in this article be concerned with the nature of an interaction. It essentially determines the way in which particles are created or annihilated. Thus in β-decay it determines, among other things, the way in which the electron and neutrino spins are oriented with respect to one another when they are emitted. In the nucleon-pion interaction it determines the orbital angular momentum with which the nucleon and pion are created and so on. Such fine details are not important for our considerations however. On the other hand it is important to consider the strength of an interaction. The strength is directly related to the *a priori* probability of the annihilation and creation process taking place and is measured by the *coupling constant*, a quantity by which the mathematical expression for the interaction is multiplied. We shall not attempt to give any quantitative representation of the interaction and it is sufficient for our purposes to note that for every process considered there is an intrinsic probability for it to take place governed primarily by the magnitude of the coupling constant.

Experimentally it is found that the interactions occurring in nature can be classified into three groups according to the values of the coupling constants. There are the *strong* interactions (coupling constant G) responsible, among other things, for the nucleon-pion interactions; the *electromagnetic* interactions (coupling constant e, where e is the fundamental unit of charge) and the *weak* interactions (coupling constant g) responsible for β-decay, decay of the pion (processes 4–6) etc. It is found that in appropriate units, G^2, e^2 and g^2 are in the approximate proportions $1 : 1/137 : 10^{-13}$. In addition to interactions that can be classified in this way there is also the gravitational interaction which, relative to these interactions, is extremely weak. However, at the present time, it seems most unlikely that it can be included within the framework being described here and it will only be briefly mentioned at the end of this article.

4. Diagrammatic representation

It has been found very convenient to represent these processes by simple pictures known as Feynman diagrams. Thus the β-decay of a neutron can be represented as in diagram 1(*a*). Diagram 1(*b*) represents the annihilation of a neutron and the subsequent creation of a proton and a π^-. 1(*c*) represents the annihilation of a neutron and a π^+ and the subsequent creation of a proton.

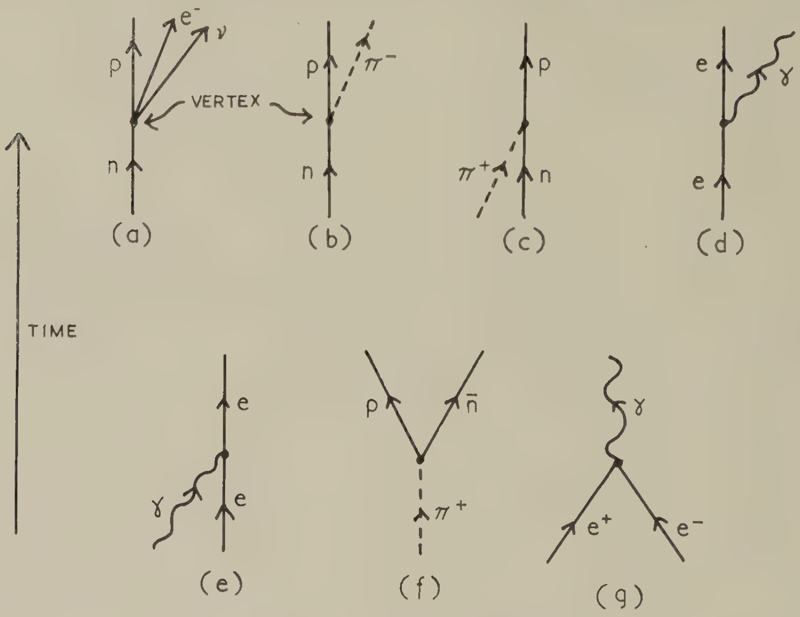


Fig. 1. Feynman Diagrams.

1(*d*) represents the emission of a photon by an electron and 1(*e*) the absorption of a photon by an electron. 1(*f*) represents the annihilation of a π^+ and the subsequent creation of a proton-antineutron pair whilst 1(*g*) represents the annihilation of an electron-positron pair with the creation of a photon. The convention with such diagrams is that time is in the direction indicated by the arrow (upwards), thus making clear whether any particular particle is being annihilated or created. The annihilation and creation is symbolically supposed to take place at the point of intersection of the various particle lines—the *vertex*. It is also the convention to represent fermions (particles of half-integer spin) by full straight lines, bosons (particles of integer spin) other than photons by dashed lines and photons (γ) by wavy lines.

For the purposes of this article these diagrams are to be regarded as purely pictorial representations of the processes under consideration. It should be mentioned, however, that such diagrams also considerably facilitate and elucidate the associated quantitative calculations related to the processes.

5. Virtual processes

In Section 2 a number of illustrative examples were discussed of processes in which particles were annihilated or created. It was stressed there that such processes could only take place if energy was conserved and that a proton, for example, cannot throw off a pion when in isolation; energy has to be provided in some way.

Some care must be taken, however, in discussing the meaning of energy conservation in such examples. The principle of conservation of energy is a principle which has been well verified experimentally. Of course, it cannot be more accurately verified than the accuracy with an energy measurement can be made. Now in our discussion we are dealing with microscopic entities and the relevant measurements are controlled by the requirements of quantum mechanics. In particular we must take due account of the restrictions imposed by Heisenberg's uncertainty relation.

This relation is one which is fundamental to the ideas of quantum mechanics and which sets an upper limit on the accuracy with which any measurement can be made quite apart from any limitations resulting from imperfections or crudities of the measuring apparatus. It can be expressed in terms of different variables, but always taken the simple mathematical form

$$\Delta \mathrm{A} \times \Delta \mathrm{B} \geqslant \hbar \tag{14}$$

The interpretation being that if two simultaneous measurements are made on a physical system of the quantities A and B, then the accuracy with which these measurements can be made (ΔA and ΔB respectively) is limited by the relation (14) where $\hbar$ is Planck's quantum constant divided by 2π ($\hbar = 1{\cdot}054 \times 10^{-27}$ erg. secs). A and B cannot be any two variables, they are restricted to certain (canonically conjugate) pairs, e.g. A = momentum in a given direction, B = position in the same direction; A = angular momentum, B = angular position; A = energy, B = time, etc. It is this last pair of quantities which is relevant to our present considerations.

The uncertainty relation for this particular case we write explicitly

$$\Delta E \times \Delta t \geqslant \hbar \tag{15}$$

where E is the measured energy of a system and t is the time at which the measurement is made. If we know this time exactly (i.e. there is no uncertainty in the time, $\Delta t = 0$) then $\Delta E = \infty$. This means that there is infinite uncertainty in the energy, or in other words, we can obtain no information at all about it. At the other extreme, we can measure the energy exactly ($\Delta E = 0$) only at the expense of not knowing at all when the energy had this measured value ($\Delta t = \infty$). Then there are intermediate situations when we can know both E and t approximately subject to the limitations of expression (15). Thus, within a time interval Δt, it is *impossible* to know the energy of a system with an accuracy greater than

$$\Delta E = \hbar / \Delta t \tag{16}$$

Under these circumstances it is not sensible or physically meaningful to require that energy be conserved to an accuracy greater than ΔE where ΔE is the upper limit on the accuracy of an energy measurement set by the uncertainty relation since such a breakdown in energy conservation could not be measured or detected *in any way*. This conclusion has considerable bearing on our discussion of the various annihilation and creation processes considered earlier.

It implies, for instance, that a nucleon can emit a pion and not conserve energy (to an extent ΔE, say) provided it does it in a time $\Delta t \leqslant \hbar / \Delta E$ and in such a way that energy is conserved *after* the time Δt. This again forbids the direct physical process

$$\text{nucleon} \rightarrow \text{nucleon} + \text{pion}$$

in the absence of any external agency since from mass considerations alone energy cannot be conserved. However it does allow a nucleon to emit (create) a pion and then subsequently reabsorb (annihilate) it, provided the whole process takes place within a time $\Delta t \leqslant \hbar/\Delta E$ where ΔE is the amount by which energy is not conserved in the emission and absorption process. The scheme of things in this case is a time sequence of two of the basic processes (13) completed in the time Δt, i.e.

$$\text{nucleon} \rightarrow \text{nucleon} + \text{pion} \rightarrow \text{nucleon}.$$

A process such as this is said to be *virtual.* For this particular case ΔE is of the order of the energy needed to create the mass energy of a pion, namely $m_\pi c^2 = 2{\cdot}24 \times 10^{-4}$ ergs, so that $\Delta t \approx 5 \times 10^{-24}$ secs. Thus emission of a pion and its subsequent reabsorption by a nucleon *can* take place provided the process is completed in a time less than approximately 5×10^{-24} secs. This sort of process is usually called a self-energy process (q.v.) and is represented diagrammatically in fig. 2. This is not the only possibility, however. Suppose there are two nucleons adjacent to each other. One nucleon could then emit a pion and the other absorb it, provided the whole process is completed in a sufficiently short time as required by the uncertainty relation and the energies of the two nucleons before and after the pion exchange are the same. The diagrammatic representation of this process is shown in fig. 3 and it is at once obvious that it provides for the interaction of one nucleon with another. This interaction may be crudely likened to that between two tennis players via the intermediary of the tennis ball.

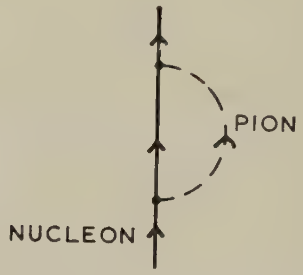


Fig. 2. Virtual emission and reabsorption of a π-meson by a nucleon.

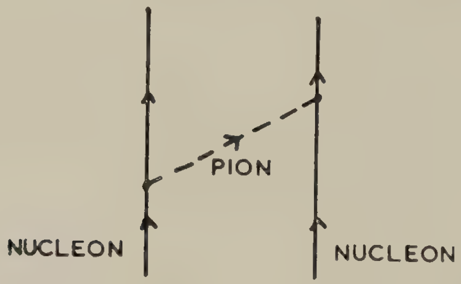


Fig. 3. Virtual exchange of a π-meson between two nucleons.

The examples we have considered here all involve nucleons and pions; however, we could also have the virtual emission and absorption of photons by charged particles, of nucleon-antinucleon pairs by pions, of electron-positron pairs by photons and so on. All such possibilities result in self energy processes and interactions between the various particles involved. We shall now consider the effects of some of these processes in more detail.

6. PROPERTIES OF THE NUCLEON

(*a*) *Size and electromagnetic properties*

To illustrate some of the previous ideas and to formulate them a little more quantitatively we investigate some of the properties they imply for an isolated nucleon. According to these ideas such a nucleon will emit and re-absorb pions during a very short time interval $\Delta t \lesssim \hbar/\Delta E$ where ΔE is the energy which has to be provided for the emission and absorption to take place. As remarked previously this energy is of the order of that associated with the rest mass of a pion, namely, $m_\pi c^2$. Thus

$$\Delta t \lesssim \hbar/m_\pi c^2 \tag{17}$$

The interesting question arises now as to how far the pion is able to get away from the nucleon before it has to return in order to get back in time to conform to the limitations of the uncertainty relation. From relativistic considerations we know that the maximum speed it can have is c, the speed of light. Thus an upper limit on this distance is of the order $r_0 = c\Delta t$ and using (17) this gives

$$r_0 \approx \hbar/m_\pi c = 1{\cdot}4 \times 10^{-13}\ \text{cm} \tag{18}$$

This is the so-called Compton wavelength of the pion.

The implication is then that a nucleon is an entity which is continually emitting and absorbing pions out to a distance of the order $1{\cdot}4 \times 10^{-13}$ cm. Physically we can therefore regard the nucleon as possessing a spherical cloud of pions of radius approximately 10^{-13} cm. In the absence of these virtual emission and absorption processes, the nucleon, pion, etc., are generally regarded as point particles. However, these processes are *never* absent; they are a property of nature and cannot be switched on and off at will. Thus the real physical nucleon as observed in nature always has its associated cloud of pions.

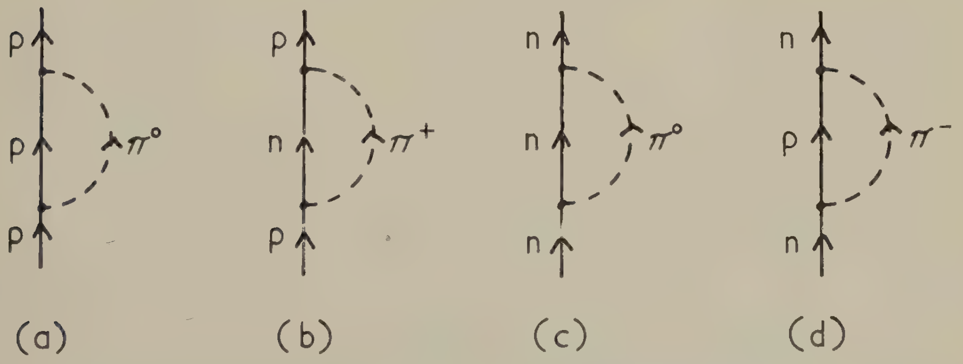


Fig. 4. Self energy processes for a proton and a neutron.

In the case of a proton, two processes are possible as shown in figs. 4*a* and 4*b*; similarly for a neutron we can have the processes 4*c* and 4*d*. In both cases, since charged pions are emitted and absorbed it is to be expected that the charges and magnetic moments of the neutron and proton should not be localized at a point but should have a finite spatial extension of about $1{\cdot}4\times10^{-13}$ cm. It is gratifying that this property has been well established by Hofstadter and his collaborators at Stanford University by measuring the way that very high energy electrons are scattered by neutrons and protons. This scattering is dependent on the nature of the electromagnetic distributions of charge and magnetic moment and his results have given very detailed information about their exact structures.

Another point to be noted is that since charged mesons are involved the actual magnitudes of the magnetic moments of the neutron and proton should differ from the values to be expected for a particle not surrounded by a pion cloud. As mentioned above such a (bare) particle does not exist in nature but we can, nevertheless, calculate its properties. This expectation is borne out by experiment. For bare particles theory gives

$$\begin{aligned}\mu_{\text{proton}} &= 1\mu_N \\ \mu_{\text{neutron}} &= 0\end{aligned} \tag{19}$$

where μ is the symbol for magnetic moment and the nuclear magneton (μ_N) is an appropriate unit of magnetic moment given by $\mu_N = e\hbar/2Mc$ with M the mass of a proton. On the other hand it is found experimentally that

$$\begin{aligned}\mu_{\text{proton}} &= (1+1{\cdot}8)\mu_N \\ \mu_{\text{neutron}} &= -1{\cdot}9\mu_N.\end{aligned} \tag{20}$$

The numbers 1·8 and −1·9 are referred to as the anomalous moments of the neutron and proton and are attributed to the effects of the pion cloud.

In the above description the situation has been presented in its simplest form and only the most elementary emission and absorption processes have been mentioned. There will, however, be much more complicated processes in which more than one pion is emitted and in which neutron-antiproton pairs may be created. A typical diagram corresponding to this situation is shown in fig. 5. Here, following the time sequence of events, a neutron emits a π^- and a proton; the π^- is then annihilated to form a neutron-antiproton pair; meanwhile the proton emits a π^0, after which the pair annihilates to form a π^- which

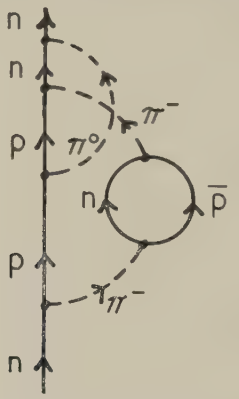


Fig. 5. Sixth order self-energy process for a neutron.

is absorbed by the proton to form a neutron, the latter finally absorbing the π^0. This complicated sequence of creation and annihilation processes must be completed in a time $\Delta t = \hbar/\Delta E$ where ΔE is the energy necessary to create the pions, neutron-antiproton pair, etc.

It is usual to speak of the *order* of a diagram by quoting the number of vertices occurring in it. Each vertex corresponds to an annihilation or creation process and associated with the process is the appropriate coupling constant G (≈ 1 in the case under discussion). In the corresponding quantitative description it is therefore not surprising that G occurs as G^n where n is the number of vertices. Thus the diagrams of fig. 4 are of order G^2 whilst the diagram of fig. 5 is of order G^6.

Now if the coupling constant is very small, the higher order diagrams and related processes are unimportant (since the probability of their taking place depends essentially on $(\textit{coupling constant})^n$) and it is a good approximation to consider only the lowest order diagrams. This situation holds for the electromagnetic interactions and the weak interactions. However, for the strong interactions, the coupling constant G is of the order unity so that all diagrams and processes are of equal importance. This means that they should all (and there is an infinity of them) be taken into account when detailed calculations are being made. Mathematically this is a formidable and at present insoluble problem and it is for this reason that although very exact calculations have been made, for example, of effects involving the virtual emission and absorption of photons it has so far proved difficult to make anything but semi-quantitative estimates of the effects generated by the strong interactions. Thus, there is at present no satisfactory calculation of the anomalous moments of the neutron and proton. It should be mentioned, however, that in the last few years the development of a new technique (dispersion theory) promises to get round much of the difficulty associated with the large value of G.

(*b*) *Self energy and renormalization*

Another important effect resulting from the virtual emission and absorption of pions by a nucleon is to modify its energy. Qualitatively it is to be expected that such an effect should arise since energy changes are clearly associated with processes involving the creation of massive particles and their subsequent absorption. Further, since no limit occurs naturally for the maximum energy with which a pion is emitted it is, perhaps, not surprising that when this 'self-energy' δE is calculated it turns out to be infinite.† For an isolated nucleon this energy can only contribute as mass energy according to the relation

$$\delta E = \delta M c^2 \tag{21}$$

Thus the observed mass of a real physical nucleon (i.e. surrounded by its pion cloud) is given by

$$M = M_0 + \delta M \tag{22}$$

where M_0 is the mass of a bare nucleon, that is the mass a nucleon would have if in some way we were able to prevent it emitting and absorbing pions. Of course, as stressed earlier, we cannot tamper with nature in this way and switch off interactions, so that M_0 is never measurable. M is therefore the sum of an unobservable mass M_0 and a calculated mass δM which is infinite.

---

† Infinities of this type are also met with classically. An example is the electrostatic energy of a point charge e. For a uniform spherical distribution of charge of radius r, the energy is $6/5\ e^2/r$ and as r tends to zero, this energy becomes infinite.

This is a curious and unsatisfactory situation. It was Kramers who first showed a way out of the apparent impasse. He pointed out that in any calculation, M_0 and δM always occur in the above combination $(M_0 + \delta M)$ and that this expression could therefore *always* be replaced by the measured physical mass M of a nucleon. At the end of a calculation only M (which is known from experiment) would appear and $M_0 + \delta M$ would drop out. This process, referred to as ‘ mass renormalization ’, has proved to be extremely successful and, particularly in quantum electrodynamics, has allowed complete agreement between theory and experiment. It smacks of trickery, but is nevertheless mathematically well defined and in spite of its intuitively unsatisfactory features is at the present time the best that can be done with a difficult problem.

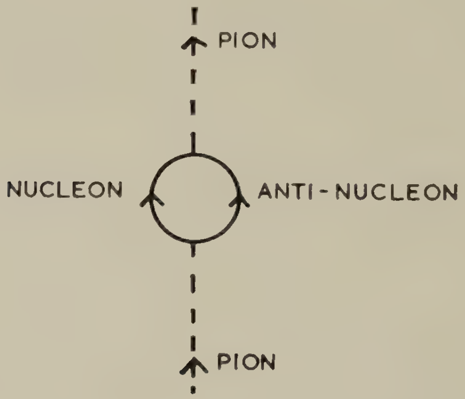


Fig. 6. Self energy process for a pion.

A pion in isolation also has an infinite self energy through the virtual creation and annihilation of a nucleon-anti-nucleon pair as in fig. 6. Again, this infinity can be removed by the process of mass renormalization, but in this case in terms of the mass of the real physical pion. However, if the pion is interacting with a nucleon and a calculation is made of the effect of a virtual nucleon-antinucleon pair (e.g. as in fig. 5) there is an additional complication in that a term appears in which the coupling constant G is multiplied by an infinite constant A (say) in such a way that the *effective* nucleon-pion interaction now has strength $G(1+A)$ rather than G. Here again G *always* occurs in combination with the factor $(1+A)$ and a ‘ coupling constant renormalization ’ procedure is used in which $G(1+A)$ is identified with the experimentally measured coupling constant G. As with mass renormalization, the main justification for this intuitively unsatisfactory procedure is that it works!

Recapitulating then we see that starting out with a theory of point nucleons and pions able to create and annihilate according to a definite scheme we finish up with physical nucleons and pions having a finite spatial extension and with, for example, considerably modified electromagnetic properties. Difficulties are encountered due to the occurrence of infite quantities in the associated calculations but these can be removed by the processes of mass and coupling constant renormalization.

7. THE INTERACTION BETWEEN TWO NUCLEONS

We now investigate some features of the way in which two nucleons can interact with one another by the exchange of pions. In just the same way that there is a self energy associated with the virtual emission of pions by a nucleon, so an energy is associated with the virtual exchange of pions by two nucleons. This energy will depend on the distance apart of the nucleons. Thus, if they are separated by an infinite distance, with whatever speed the pions are emitted by one nucleon it would take an infinite time to reach the other and so energy could not be conserved within the time set by the uncertainty relation (15). The process, therefore, cannot take place and consequently the interaction energy is zero for infinite separation. We have seen, however, that a pion can be emitted up to a distance of approximately $1{\cdot}4\times 10^{-13}$ cm. from a nucleon without violating the principle of energy conservation. This implies that if two nucleons are within this distance of one another they can interact through pion exchange and there will be an associated interaction energy dependent on the exact distance apart. We denote this energy by $V(r)$ and identify it with the internucleon potential energy in terms of which the internucleon force can be described (eqn. (3)). Here r is the distance between the nucleons and from our qualitative considerations we know that

$$\begin{aligned} V(\infty) &= 0 \\ V(r) &= \text{finite for } r \lesssim 1{\cdot}4\times 10^{-13}\text{ cm.} \end{aligned} \tag{23}$$

Detailed calculations show that the actual dependence of $V(r)$ on r is controlled primarily by a factor $\exp(-r/r_0)/r$ where r_0 is the Compton wavelength of the pion ($\hbar/mc$) given in eqn. (18). This dependence on r was first deduced by Yukawa and this type of potential is usually referred to as a Yukawa potential. Its r dependence satisfies the above qualitative requirements; it vanishes for $r \to \infty$ and dies away quickly for $r > r_0$. For this latter reason r_0 is called the ‘ range ’ of the potential.

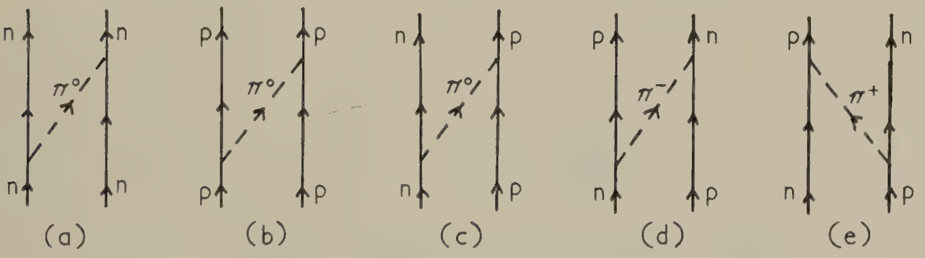


Fig. 7. Diagrams contributing to the internucleon potential.

The simplest types of diagram contributing to the interaction between two neutrons, two protons and a neutron and a proton are shown in fig. 7. In each diagram there are two vertices with associated coupling constant G and so it is not surprising that if the nucleons are stationary the potential energy calculated for each diagram has in each case the essential form

$$V(r) = -G^2 \exp(-r/r_0)/r \tag{24}$$

In addition to the diagrams of fig. 7 there will also be contributions from higher order diagrams in which more than one meson is exchanged and, since G is of the order of magnitude unity, they are equally important. Their range, however, is expected to be smaller because of the greater energy in the virtual states. Qualitatively speaking, to create for example, two pions, more energy ($2m_{\pi}c^2$) is needed so that the uncertainty relation allows a shorter time ($\approx \hbar/2m_{\pi}c^2$) for the pion exchange to take place and hence only a shorter distance ($\hbar/2m_{\pi}c$) can be traversed by the virtual pions. This qualitative result is in agreement with more quantitative estimates. Thus the long range properties of the potential are determined primarily by the diagrams of fig. 7; more complicated diagrams contribute mainly to the short range properties.

A feature of the internucleon potential which is not met with classically is that it possesses an ' exchange ' character. By this is meant that when, for example, a neutron and a proton interact with one another they can exchange their rôles. In a collision between a neutron and a proton in which scattering is brought about by the internucleon potential, they can scatter in such a way that a neutron appears where the proton might be expected and vice versa. The diagrams 7*d* and 7*e* are responsible for this type of effect.

In the foregoing discussion we have completely ignored the fact that the nucleons possess spin ($\frac{1}{2}\hbar$) and may also be moving so that in general it is to be expected that $V(r)$ should also depend on their relative spin orientations and relative momentum. Theoretical and experimental investigation shows that there is indeed a very strong spin dependence manifested among other things through the ' tensor ' part of $V(r)$. Further $V(r)$ also depends on the orientation of the spins with respect to the relative orbital angular momentum of the two nucleons; this is referred to as ' spin-orbit coupling '. Finally, although $V(r)$ is predominantly an attractive potential (as it must be to bind nucleons together to form stable nuclei), for $r \lesssim 0{\cdot}4 \times 10^{-13}$ cm it becomes highly repulsive. It is difficult to calculate $V(r)$ for such a small value of r since many virtual pions will be contributing and, as mentioned earlier, there is at present no mathematical technique for summing the large number of diagrams that must be contributing. This last property of $V(r)$ is one which is deduced primarily from experimental investigations of nucleon-nucleon scattering. The relevance of all these properties of $V(r)$ to nuclear structure is discussed in a previous article$^{(2)}$ by the author.

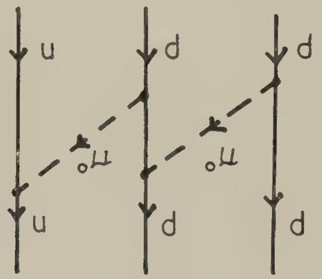


Fig. 8. Diagram contributing to the three-body potential.

The discussion in this section has been restricted to the case of two nucleons interacting with one another. For completeness it should be pointed out that the interpretation of the internucleon potential in terms of pion exchange also allows and, in fact, predicts the existence of many body forces, that is, forces which only contribute when more than two nucleons are adjacent to one another. For instance a three-body potential can arise from the diagram of fig. 8 and it is straightforward to construct diagrams involving higher numbers of nucleons. Such many body potentials are not believed to be of any great importance. The most likely place for them to manifest themselves is in nuclear structure where we are dealing with many nucleons in close proximity to one another. Even here, however, the chance of three (say) nucleons to be *simultaneously* close enough to each other to experience such a potential is very small and so far two-body potentials alone have sufficed to understand and interpret nuclear properties.

8. ELECTROMAGNETIC EFFECTS

We now leave the nucleon-pion interaction and go on to consider the effects of the virtual emission and absorption of photons by charged particles.† Here it is most useful to focus our attention on the interaction of electrons with photons, since electrons do not experience any strong interactions which would mask the weaker (by a factor ≈ 137) electromagnetic interactions.

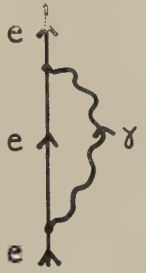


Fig. 9. Electromagnetic self energy process for an electron.

An electron in isolation experiences self-energy processes due to the virtual emission and absorption of photons of the type shown in fig. 9. As with the analogous situation arising with the nucleon-pion interaction the calculated self-energy here is infinite, but can again be removed by the process of mass renormalization. Similarly the virtual creation and annihilation of electron-positron pairs by a photon does not lead to a finite mass for the photon.

Consider next the interaction of an electron with an external magnetic field; such an interaction is relevant to any measurement of the electron magnetic moment. The basic interaction is denoted symbolically by diagram (*a*) of fig. 10 in which the cross and associated photon line is a symbolistic representation of the external field interacting with the electron. There will be additional effects of order e^2 (e being the relevant coupling constant in this case) due to (*b*) virtual creation and annihilation of an electron-positron pair (referred to as vacuum polarization), (*c*), (*d*) and (*e*) virtual emission and absorption of a photon by the electron. As in the nucleon-pion interaction all these diagrams lead to infinities which can be removed by mass and (in this case) charge renormalization. There is a slight difference in the present case, however, namely, that the electron is not isolated ; it is interacting with the magnetic field. Under these circumstances it is not surprising that the infinities† which arise differ a little from those which arise in the case of a free electron. Thus, even when mass and charge renormalization is applied in a calculation of the energies associated with diagrams 10(*b*)–10(*e*) there is still an energy left over proportional to e^2. This energy turns out to be finite and it arises in such a way that it can be interpreted as due to a modification of the magnetic moment of the electron over and above the value to be expected for a ‘ bare ’ electron (i.e. an electron not able to experience the virtual processes). Quantitatively, the following expression is obtained for the electron magnetic moment taking into account the diagrams of fig. 10

---

†This branch of field theory is usually referred to as quantum electrodynamics.

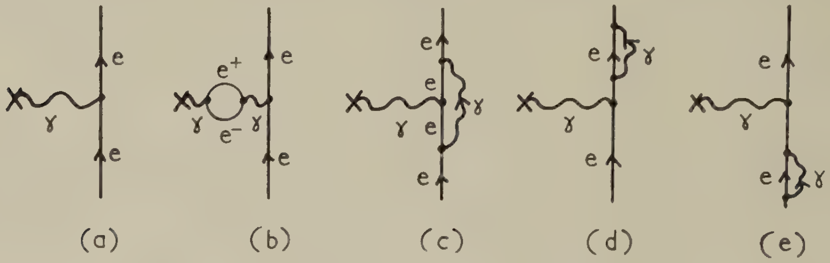


Fig. 10. (*a*) Interaction of an electron with an external field (*b*)–(*e*) Inclusion of virtual processes.

$$\mu_{\text{electron}} = \mu_B(1 + e^2/2\pi\hbar c) \tag{25}$$

where μ_B is the Bohr magneton and is the magnetic moment predicted (by Dirac) for a bare electron. Explicitly $\mu_B = e\hbar/2mc$ where m is the electron mass. (25) gives

$$\mu_{\text{electron}} = 1{\cdot}0011614\ \mu_B$$

whereas experimentally

$$\mu_{\text{electron}} = (1{\cdot}001165 \pm 0{\cdot}000011)\ \mu_B.$$

The small discrepancy can be fully accounted for by considering the effects of higher order diagrams.

Another success of quantum electrodynamics has been in accounting for the Lamb shift in the hydrogen atom. If calculations are made using ordinary relativistic quantum mechanics (that is, not taking account of virtual processes) of the electron energy levels in the hydrogen atom it is found that two levels, $2s_{\frac{1}{2}}$ and $2p_{\frac{1}{2}}$ have *exactly* the same energy. Here the labelling of states is as usual (cf. reference 1); the figure 2 refers to the radial quantum number, s and p indicate that the *orbital* angular momenta in the two states are respectively 0 and $\hbar$ and the $\frac{1}{2}$ implies that the *total* angular momentum in each state is $\frac{1}{2}\hbar$. Experimentally it is found that the levels are shifted (Lamb shift) relative to one another by a small amount ΔE where in terms of the corresponding frequency $\Delta\nu_{\text{exp}} (\Delta E = h\Delta\nu_{\text{exp}})$ of an electromagnetic transition between the two levels

$$\Delta\nu_{\text{exp}} = (1057{\cdot}77 \pm 0{\cdot}10)\ \text{Mc/s}.$$

This shift can easily be accounted for in much the same way as the small change in the magnetic moment of the electron. In fact, to second order, we can use the diagrams of fig. 10, except that the external field is now no longer a magnetic field but is the Coulomb field of the hydrogen nucleus (proton). Again infinities arise in calculating the energies and when these are removed by the renormalization procedure there is still a finite, albeit small, energy left over. This energy is different for the $2s_{\frac{1}{2}}$ level and the $2p_{\frac{1}{2}}$ level since, having different orbital angular momenta, the spatial motion of the electron in these states are different. Since these energy changes due to virtual processes differ for the two states it follows that the $2s_{\frac{1}{2}}$ and $2p_{\frac{1}{2}}$ levels no longer coincide. To order e^2 it is found that the corresponding frequency difference $\Delta\nu_{\text{theoretical}}$ is

---

† The infinities that occur in field theory are all well defined mathematically as limiting values of certain integrals.

$$\Delta\nu_{\text{theoretical}} = 1052{\cdot}14 \text{ Mc/s}.$$

The small discrepancy between theory and experiment can again be removed if higher order diagrams are taken into consideration.

These remarkable agreements between theory and experiment represent a great triumph for quantum electrodynamics and lead one to believe in its essential correctness in spite of the artificiality of the renormalization procedures. It should be stressed again that this agreement is only possible because of the smallness of $e^2/\hbar c$ $(=1/137)$ and the consequent result that only a few low order (small number of vertices) diagrams have to be taken into account in making the calculations.

One final point to be considered is the nature of the potential between two charged particles. We saw in the case of the interaction between two stationary nucleons that the potential had the Yukawa form (eqn. 21), namely,

$$V(r) = -G^2 \exp(-r/r_0)/r$$

where $r_0 = \hbar/m_\pi c$. In the electromagnetic case the potential is propagated by a photon of mass zero rather than a pion of mass m_π. Thus r_0 is infinite. Further the coupling constant is e rather than G so that the potential between two stationary charged particles should now have the above form but with $G \to e$ and $r_0 \to \infty$, that is

$$V(r)_{\text{electric}} \sim e^2/r.$$

This is, of course, the well-known Coulomb potential. The long range character of this potential stems from the fact that the potential results from the exchange of a massless particle, the photon.

Quantum electrodynamics holds a unique place among the various field theories because of the zero mass of the photon and at the present time all experimentally observed electromagnetic properties are fully accounted for by the theory.

9. DISCUSSION

In the foregoing sections we have only considered the nucleon-pion and electromagnetic interactions. There are, of course, many others, in particular those involving the so-called strange particles⁽¹⁾. The strange fermions (Λ, $\Sigma^\pm$, Σ^0, Ξ^-, Ξ^0) all have masses greater than the nucleon (N) mass and the strange bosons ($K^\pm$, K^0, $\bar{K}^0$) have masses $m_K \approx 960\, m$, where m is the electron mass. These particles interact strongly (with a coupling constant $\approx G$)

according to schemes of the type:

$$\begin{aligned}\Lambda &\rightleftharpoons \Sigma + \pi \\ \Lambda &\rightleftharpoons \mathrm{N} + \mathrm{K} \\ \Sigma &\rightleftharpoons \mathrm{N} + \mathrm{K} \\ \Xi &\rightleftharpoons \mathrm{N} + \mathrm{K} \\ \Sigma &\rightleftharpoons \Sigma + \pi\end{aligned} \tag{26}$$

with appropriate interchanges of particles to antiparticles, etc., as in the schemes (10). It is to be noted, however, that the interaction $\Lambda \rightleftharpoons \Lambda + \pi^0$ does not occur.

The details of the various interactions are not fully understood and at the present time only very approximate calculations of the properties and interactions of the particles have been made. As an example brief mention might be made of the potential existing between a Λ-particle and a nucleon. This is of particular interest since there is a lot of experimental data available on hyperfragments (or hypernuclei). Hyperfragments are light nuclei in which one of the neutrons is replaced by a Λ-particle. Eventually (within $\approx 10^{-10}$ sec) a hyperfragment decays since a Λ-particle is unstable against the (weak) decay processes.

$$\Lambda \begin{matrix} \rightarrow n + \pi^0 \\ \rightarrow p + \pi^-. \end{matrix}$$

Nevertheless reliable measurements can be made of the binding energy of a Λ-particle in a hyperfragment and properties of the Λ-nucleon potential thereby deduced.

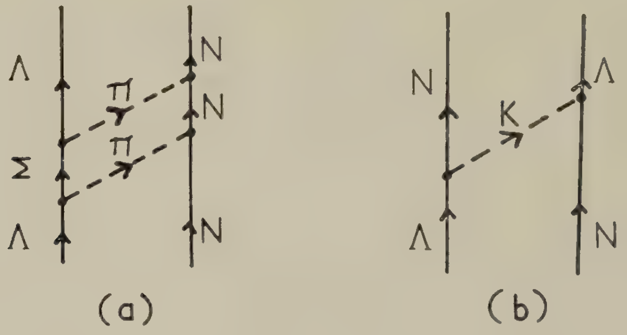


Fig. 11. Diagrams contributing to the Λ-nucleon potential.

In terms of the interactions (23) the Λ-nucleon potential can arise, for example, from diagrams of the type shown in fig. 11. In 11(*a*) two pions are exchanged and in 11(*b*) one *K* meson. The former potential then has a Yukawa-like dependence with range $r_0 = \hbar/2m_\pi c$ and the latter has a range $= r_0 \hbar/m_K c$ and also has an exchange character (see section 7). Since $2m_\pi \approx 550\ m$ and $m_K \approx 960\ m$, both of these potentials have a range shorter than that of the inter-nucleon potential ($r_0 = \hbar/m_\pi c$). Calculations seem to indicate that the former is the most important and that its properties account for the main features of Λ-particle binding in hyperfragments.

The weak interactions responsible, among other things, for β-decay, decay of the π-meson and μ-meson and of the strange particles have also received a lot of attention during the last few years. Here the relevant coupling constant g is very small so that only the lowest order diagrams have to be considered. Difficulties are still encountered, however, because of the modifying effects on the interacting particles of the strong interactions and one is again faced with the problem of giving a proper treatment of diagrams involving the strong interactions to high orders. Another formal difficulty also arises in that, as usually formulated, the weak interactions are not renormalizable. Although the coupling constant g is very small, when higher order weak interaction diagrams are calculated they contribute infinities multiplied by some power of g. However small g is the result is still infinite and unfortunately it occurs in such a way that mass and coupling constant renormalization procedures cannot be used. At present this difficulty is usually put on one side and ignored and only the lowest order diagram, which is finite, for a particular process is considered.

A considerable amount of effort is being expended at present in trying to incorporate the interactions we have been considering (4), (5), (6), (13) and (26) into some universal scheme in which all the particles involved occur in some symmetrical fashion. Such phrases as ' Global Symmetry ', ' Universal Fermi Interaction ', etc., commonly occur in the titles of many published papers on this subject. So far, however, no universal scheme has been found which satisfies all the requirements of the observed physical facts. In such schemes the particles are introduced in an *ad hoc* fashion and no consideration is given as to the ' why ' or ' how ' the particles occur with their observed properties.

Another approach to this problem has been to build up composite models of particles. Thus interaction (9), for example, might imply that a neutron is in some sense a bound state of a proton and a π^- (in much the same way that a hydrogen atom is a bound state of a proton and an electron) or conversely that a π^- is a bound state of a neutron and an anti-proton. Similarly, interactions (26) suggest that a Λ-particle may be a bound state of a neutron and a K_0 meson and so on. By such a description it is possible to reduce the number of *really* elementary particles. Specific schemes of this type have been suggested but nothing really satisfactory emerges. The ultimate of such a scheme is currently being investigated by Heisenberg who proposes that there is really only *one* ' field ', a matter field not to be identified with any of the known particles, which interacts with itself in such a way that each elementary particle arises as some sort of singularity in the field. The theory would thus predict the masses, spins, etc., of the elementary particles. Unfortunately the mathematical difficulties and the interpretation of such a scheme are prodigious and so far little headway has been made with it. It has, however, considerable aesthetic appeal.

Finally a remark is in order about the gravitational interaction. It seems at present difficult to group this with the interactions we have discussed, although attempts have been made, and to account for it in terms of the exchange of some hypothetical particle. Rather it seems more natural to associate the phenomenon of gravitation with what might be called the ' geometry ' of space. It is bound up with such fundamental problems as the nature and evolution of the universe but such problems do not properly fall within the scope of this article.

10. Conclusion

The aim of this article has been to try and give some account of the physical ideas underlying quantum field theory, a theory which enables a description to be given, albeit sometimes only qualitative, of the way in which elementary particles acquire their various properties and interact with one another. The interactions and forces existing between macroscopic pieces of matter then result as the sum total of these basic elemental interactions. In attempting to achieve this aim accuracy has frequently been sacrificed at the altar of pictorial clarity. Nevertheless, the scheme as presented is an approximate representation of what can only be properly expressed in terms of extremely complicated mathematics. In that such a representation glosses over difficulties and misrepresents to some extent it is open to criticism from the purist, and there are many such working in this particular field. It is this author's contention, however, that if a glimmering of the field theorist's activities can be communicated to the general reader, even at the expense of inaccuracy, then it is well worth the purist's scorn. There is far too little communication between different specialists today, let alone specialist and layman, and every possible attempt should be made to break down the not impenetrable barriers between them.

References

(1) Salam, A., *Contemporary Physics*, **1**, 337.
(2) Blin-Stoyle, R. J., 1959, *Contemporary Physics*, **1**, 17.

Reading list

To the author's knowledge there are no really elementary accounts of quantum field theory. It might be of some profit to the reader who has a working knowledge of ordinary quantum mechanics to refer to the following two volumes:

Elementary Particles, by Enrico Fermi (New Haven: Yale University Press, 1951).
Introduction to Quantum Field Theory, by F. Mandl (London : Interscience Publishers Ltd., 1959).

The Author:

Dr. Roger Blin-Stoyle was educated at Alderman Newton's Boys' School, Leicester, and at Wadham College, Oxford. He spent a year as Lecturer in Mathematical Physics at Birmingham University and then returned to Oxford in 1954 as Senior Research Officer in Theoretical Physics. His research has been in the field of theoretical nuclear physics and elementary particle physics. He was elected a Fellow of Wadham College in 1956.

Electrification Following the Contact of Solids

by W. R. HARPER

SUMMARY

The charge found on solids after they have been rubbed together can only be a direct consequence of friction when the rubbing is vigorous; more usually, the charging is the result of contact, rubbing complicating the phenomenon. Opinions differ regarding the origin of contact charging, but certain experiments suggest a working hypothesis, according to which the charging is due to electron transfer when both materials are either semiconductors or metals, but to ions when insulators take part. Electron transfer is associated with the setting up of a contact potential. Ion transfer can be caused either by a potential difference, in which case the current flows with the e.m.f., or it can be caused by forces that are not electrical in nature, in which case the current flows against the potential difference generated by the 'mechanical transfer' of charge.

1. STATIC ELECTRIFICATION

The facility with which different materials acquire a static electric charge depends on how we assess their behaviour. The ability of amber to attract light objects after being rubbed was only noticed long ago because the charge is retained for at least some seconds: amber is a good insulator. So few good insulators were known in antiquity that the ability to become electrified must have seemed highly peculiar. It was found later that improving the surface insulation of such materials as glass by drying, and at the same time using a simple electroscope to investigate the charging, led to the conclusion that all common insulators can be charged. It was furthermore found that conductors also could be charged, provided they were insulated from earth, so that the charge could not leak away. Further investigation led to the belief that the mutual electrification of surfaces rubbed together was an almost universal phenomenon, so much so that, in 1917, Shaw was able to comment that "it is a commonplace in tribo-electricity to find two surfaces apparently identical which, when rubbed together, excite one another with opposite charges". Shaw was speaking of surfaces that were clean in a workshop sense, perhaps in a laboratory sense, but certainly not on a molecular scale. As we shall see later, he was really thinking of charges that must be considered as small from a theoretical standpoint, though quite easy to measure experimentally.

The production of synthetic polymers, most of which are good insulators, has provided a range of materials which, when electrified, hold their charge for a long time, for which reason they have acquired the reputation of being materials that are easily electrified. Certainly, once they become charged, they seem determined to stay that way. Careful investigation, however, has revealed that they are actually far *less* inclined to become electrified on contact than are such insulators as glass and quartz. Harper, in 1953, compared the charges acquired by different insulators after light contact with metals, and found that insulators fell into two groups. A good deal of trouble was taken to clean the insulators, but the methods used were likely to have been much more successful on chemically inert insulators such as synthetic polymers, than on others such as quartz. Many insulators charged as had been expected, but polythene, polystyrene and a number of others showed only negligible charging by comparison, of the order of 1 per cent of that found with the former group. The existence of *electrophobic* insulators (those reluctant to charge) came as a great surprise, particularly since amber, from which electricity takes its name, proved to be electrophobic. The surprise was intensified in 1955 when Harper found that quartz, if cleaned with scrupulous care, showed charging that can only be described as enormous. It was around a hundredfold the amount obtained with quartz that would previously have been considered well cleaned, and *four orders of magnitude* larger than any residual charging obtained with polystyrene. Moreover the record large charging came from the contact of two surfaces, *both of which were quartz*, though different crystal faces. Here, then, is an extreme case of a second kind of insulator which has a propensity for becoming charged—an *electrophilic* insulator.

We must now ask which group of insulators conforms to type, and which is exceptional; is it the electrophobic insulators which do not charge because they are truly insulating, or is it the electrophilic insulators which charge as an insulator should, there being some obstacle which prevents the electrophobic insulators from conforming to type? Then, if the latter, is the extremely large charging of very clean quartz to be regarded as typical, or the intermediate charging of the not quite so clean insulators? It seemed at first natural to suppose that the extremes of charging, the enormous charging of quartz and the relatively negligible charging of polystyrene, were exceptional and required some special explanation, the intermediate charging being normal. Intermediate charging at any rate includes the results of most experiments that have been described in the literature. We shall, however, find there are good grounds for believing that, on the contrary, it is the extremes of charging that must be regarded as being significant theoretically, and that the electrophobic insulator behaves as a simple insulator should. On this view, contact charging ceases to be characteristic of the true insulator.

So far, no careful distinction has been made between charging resulting from contact (followed by separation), and charging due to rubbing. Volta, in 1779, taught that there is no fundamental difference, the effect of rubbing being merely to multiply the points of contact. This opinion was developed by Helmholtz in 1879, and has found great favour ever since, always admitting that friction may change the surfaces in contact, and so give rise to secondary effects. Nevertheless, what we shall call the Volta/Helmholtz hypothesis has also found opponents, and there is not sufficient justification for a proposal that has been made to call all triboelectric charging contact charging. On the other hand, the term ‘ *tribo*electric ’ is unfortunate in begging an unsettled question—probably mostly incorrectly.

‘ Frictional ’ charging is temperamental to an extent that friction is not, and experiments performed by different observers in different places not infrequently appear at first sight incompatible. It must be supposed that the charging may be determined by true surface properties which can be altered by a change in the topmost monomolecular layer, whereas friction itself usually involves a finite depth of material below the surfaces, so that the bulk properties of the substances are responsible for what is observed. It is easy to make a list of factors which may influence the results of experiments on the generation of static charge, but they are so many, and so difficult to control with certainty, that little help is to be derived from doing so. The subject is one in which clear-cut results are the exception rather than the rule, quite unlike what one prefers to regard as the characteristic of physics—precise unambiguous measurements. In these circumstances interpretation is difficult and often a matter of personal opinion; no useful overall survey of the subject is yet possible. In the present article, therefore, attention will be confined to a limited selection of experiments, chosen for discussion because they appear to make some sense against a background of modern theories of the solid state, or because they are important in seeming to contradict theoretical expectations.

2. ELECTRIFICATION DUE TO FRICTION

If the mutual rubbing of two surfaces makes one positively charged, and the other negatively, some asymmetry must exist. When the two surfaces are seemingly identical, they may really be differently contaminated, and the charging due to the contamination. If clean, the difference between the surfaces may be one of crystal anisotropy, which can be of great importance, as in Harper's experiments on quartz. Shaw thought that a difference in the state of strain could be important, but this idea has not proved a fruitful one. The asymmetry, however, need not belong to the surfaces: it may be an asymmetry of rubbing. This was demonstrated by Henry by an experiment in which it was found that the charging behaviour of two like surfaces was interchanged when the manner in which one rubbed the other was reversed. Two crossed cylinders were used, the one being moved back and forth like a saw over the other. The cylinder rubbed as if it were being sawn is rubbed at one spot, and vigorous rubbing can make this spot hot; the cylinder moved like a saw has the rubbed area drawn out to the length of the rub so that a narrow ribbon of surface is heated, and the ribbon does not get so hot. Henry's experiment was performed with amorphous highly insulating materials. The observed charging *must* have been due to friction, and almost certainly to heating; the Volta/Helmholtz hypothesis is ruled out in this case. The materials were electrophobic: for such insulators mere contact is insufficient to bring about the transfer either of electrons or of positive ions. A sufficient rise in temperature turns an insulator, for which electrons in the filled valency band have too little thermal energy to reach the conduction band, into an intrinsic semiconductor, for which a few electrons do reach the conduction band because of their increased thermal energy. This will happen to a different extent on the two sides of a contact at which asymmetric rubbing occurs, so that there are, in effect, two different semiconductors in contact. Alternatively, one may say that there is one semiconductor with a temperature difference between the two parts, and deduce that a Thomson thermoelectric potential will arise between them, and hence a charge of electrons be transferred in one direction or the other. Moreover, this explanation of the charging appears to be the only one possible if it is electrons that move across the interface, which is likely since, as a general rule, when the temperature is raised, electrons become mobile long before positive ions.

The asymmetric rubbing experiment is important in proving that *electrification can be purely frictional* in origin, and because one explanation seems much more plausible than any other. The explanation calls on the only thermoelectric e.m.f. that arises in a single material—the Thomson effect. The suggestion that a thermoelectric e.m.f. must be responsible may appear surprising: thermoelectric effects are usually thought of as small. This, however, is only true of metals; in semiconductors these potentials are often much larger, and our explanation of the charging depends on the insulator becoming a semiconductor when rubbed.

If the Thomson e.m.f. can give rise to triboelectric charging with like materials, the Seebeck e.m.f. will replace it with dissimilar materials, but will appear as a component of the contact potential. In the case of metals, the Seebeck effect only modifies the contact potential slightly, and may usually be neglected, but in the case of semiconductors (and insulators which, through heating, have temporarily become semiconductors) the effect, though secondary, may be important. In principle, there is a contact potential between all dissimilar materials. When the materials can conduct electricity, the contact potential can give rise to charging across an interface, though no continuous current will flow round a closed circuit because the contact potentials at different interfaces cancel out (apart from the Seebeck and Thomson components).

It was suggested by Frenkel that the ‘ triboelectric ’ charging of insulators is normally due to the formation of hot spots during rubbing. The insulators become intrinsic semiconductors where hot, and a contact potential is enabled to move electrons across the interface, the charge being ‘ frozen-in ’ when the hot spots subsequently cool. Owing to a numerical mistake, however, Frenkel overestimated the importance of hot spots, and revised calculations make it clear that, in many important cases of ‘ triboelectric ’ charging, hot spots cannot be significant. Loeb is of the opinion that the generation of heat in friction has seldom to do with the transfer of charge, and with this view the present author concurs. We have nevertheless seen that heating must be important in some cases; these are probably limited to vigorous rubbing.

If these views about the significance of friction are correct, the virtue of the heat generated in the hot spots lies in the electrical conductivity to which it gives rise. The work expended in friction acts indirectly through the rise in temperature, hence there is no reason to expect a correlation between the work done and the charge generated. Attempts that have been made to find such a correlation have, in fact, failed.

3. Electrification due to contact

If friction is seldom the prime cause of electrification, it must nevertheless frequently complicate the generation of charge as a result of contact. An illustration is given in fig. 1, taken from the work of Harper on the electrification of metals. The two histograms of the size of charges observed when sliding was encouraged, and when sliding was thought to have been eliminated, show how behaviour that is consistent with simple contact is made so haphazard by the introduction of sliding that even reversed charging becomes prominent. The charging in question was brought about by the contact of spheres; only a single contact ‘ point ’ being involved. When extended surfaces are used in experiments, the charge found will be the sum of a large number of individual charges at such ‘ points ’. Inconsistent behaviour may be averaged out, whence it follows that consistency among readings which are really averages gives no guarantee that the true behaviour is uniform. Many experimenters, however, have tacitly assumed that it does give such a guarantee, but the author’s experience with apparatus in which single contact was used suggests that a greal deal of complexity still remains obscured because of experimental methods employing multiple contact.

The complete elimination of sliding during the making and breaking of contact is very difficult, but nature provides one simple way of achieving it—by using mercury as one of the contacting materials. Medley has obtained very large charges from a mercury/polythene contact, so demonstrating that simple contact is, on occasion, sufficient to give rise to charges of the order of magnitude found in ‘ triboelectrification ’, *without the intervention of friction.*

The large charges from mercury/polythene contrast strangely with the finding that polythene is an electrophobic insulator, it being one that showed negligible charging in Harper’s experiments, and particularly since polythene has the reputation of charging freely in industrial practice. The duration of contact in Harper’s experiments was the order of 1/5 sec, in Medley’s experiments the order of 10 sec. It seemed possible that prolonged contact is more effective than short contact, but investigating this with Harper’s apparatus using contacts up to two hours still gave no charging. Some other explanation is required. In fact, there are two quite different considerations relevant to the charging of polythene, polystyrene and other materials which retain surface charges persistently. One consideration concerns the meaning to be attached to ‘ clean ’, the other concerns what are to be regarded as large (or small) charges.

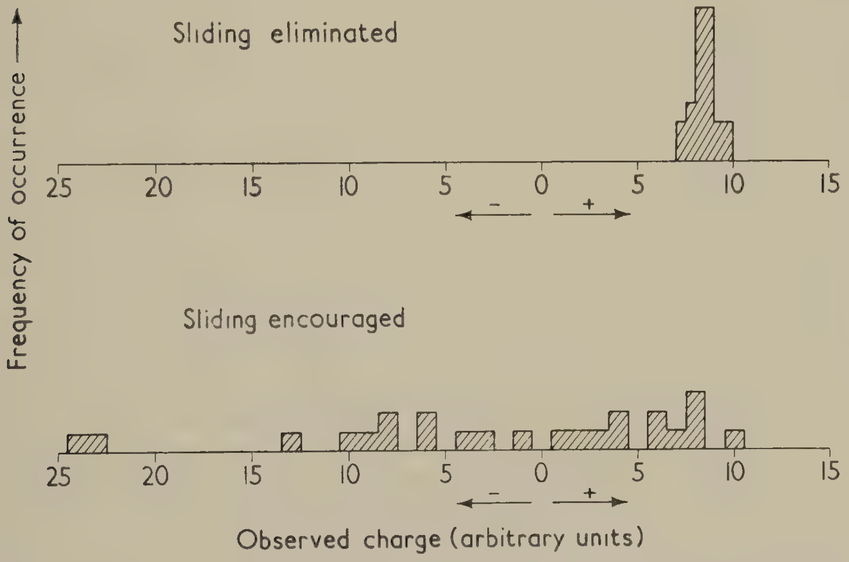


Fig. 1. Effect of sliding on charging.

When two experimenters obtain incompatible results, it is tempting for each to accuse the other of having worked with dirty surfaces—and both would probably be right. There is no such thing as a strictly clean surface in this subject. Metal surfaces can only be obtained free from the last adsorbed monomolecular layer by prolonged baking at a high temperature *in vacuo*. A crystal divided into two parts along a crystallographic plane has loose bonds left exposed where the two parts were previously bound together, and these loose bonds will promptly collect something from the surroundings in order to reach a stable condition. A surface layer of special character is formed, its nature depending on how the stabilization occurred. In general, a stable surface, even when free from unnecessary contamination, will be different from the underlying structure, and in that sense not clean. It is better to think in terms of how a surface *has been prepared*. Now it is known from experiments on the spraying of mercury that its surface charging properties are measurably affected by the presence of as little as one part of zinc in 10^{11} by volume. This can only happen if the zinc becomes concentrated in a surface skin (probably as oxide). Since no special precautions were taken by Medley to purify his mercury, some surface skin is certain to have been present. In the experiments of Harper which gave no charging with polythene against metals, these metals were undoubtedly covered with an oxide surface layer, but the layer will have been firmly attached by cohesive forces, whereas a thin skin on mercury is obviously freely detachable. The different behaviour in the two experiments could therefore be accounted for if the charging in Medley's apparatus was due to transfer of *material*, carrying charge with it. A mechanism of this kind has more than once appealed to workers in the subject as an explanation of the generation of static charge, but there is little evidence concerning its importance.

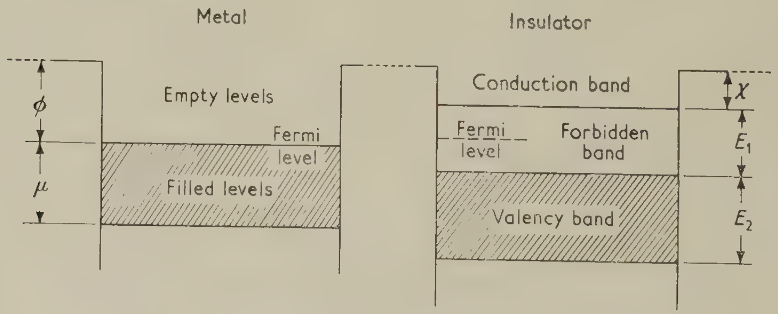


Fig. 2. Electron levels in metal and insulator.

It is difficult to say whether a similar explanation can account for the charging commonly found when handling polystyrene film and similar materials, but it must be recognized that a charge which is sufficient to have great nuisance value may nevertheless be small from the theoretical standpoint, and due to some secondary effect. To assess what charge density is to be expected from a primary cause of charging, we must have a model of the supposed mechanism. Let us consider, firstly, contact charging consequent on electrons moving to establish a contact potential, and, secondly, contact charging due to a transfer of positive ions.

Figure 2 shows the distribution of electron levels in a block of metal, and in a block of insulator. The available band of levels is almost exactly half filled in the case of the metal. The insulator is characterized by a forbidden band between the filled valency band and the conduction band (which, however, does not actually conduct because the number of electrons in it is negligible). If two metals having different work functions ϕ_1 and ϕ_2 (volts) are brought into contact, electrons from the top of the filled set of levels in the metal with the smaller work function can fall into unoccupied levels at the bottom of the empty set of levels in the other metal. By doing so they will charge it negatively, leaving an equal positive charge behind. These charges attract one another, and therefore remain close to the interface, forming a double layer. The two halves of the double layer give rise to an electric field between them, but elsewhere the two contributions to the field from the positive and negative charges cancel out. There is therefore a potential discontinuity on passing through the double layer, of amount $4\pi\sigma d$, where d is the thickness of the layer, and σ the surface density of charge. It is clear that the transfer of electrons from one metal to the other will stop when the potential discontinuity across the interface has reached $\phi_1 - \phi_2$, that is when the tops of the filled parts of the bands in the two metals are levelled up. It is this process that establishes the contact potential between the metals, by the formation of a double layer. In the case of metals the thickness of the double layer is of the order of molecular dimensions.

If a similar process were possible for insulators, and contact charging was due to the two halves of the double layer remaining each with its own insulator after separation, the resulting surface density of charge would be the contact potential divided by $4\pi d$. Contact potentials are of the order of several volts. It is, however, certain that some of the charge of the double layer would be lost during separation, because wave-mechanics permits electrons to jump a small gap by tunneling through the potential barrier. Compromising by taking a value of one volt $= 1/300$ e.s.u., and assuming $d = 10^{-8}$ cm, we find $\sigma \approx 2 \times 10^4$ e.s.u./cm^2.

Positive ions, because of their greater mass, cannot tunnel like electrons; if there is a redistribution between two surfaces in contact, the new distribution will remain when the surfaces are separated. We would not expect every atom exposed on a surface to be ionieed and to move to another surface on contact with it, but it seems quite possible that one in ten might do so. Taking a lattice square as 2×10^{-15} cm^2, the resulting surface density of charge would then be approximately 2×10^4 e.s.u./cm^2, the same as the value estimated for electron transfer.

These estimates are crude, but their interest lies in the very large value of surface density predicted. A surface density of 1 e.s.u./cm^2 is noticeable under favourable circumstances, such as when the charge resides on thin plastic film; at about 16 e.s.u./cm^2 an extended charge will begin to discharge into the atmosphere. Typical values in charging experiments range between 1 and 10 e.s.u./cm^2; with powders, charge densities may be an order of magnitude less. There are three reasons apparent why observed values should be much less than the theoretical estimate:

(i) Charge density limited by discharge to the surroundings.
(ii) Only a fraction of the surface carrying the charge contributing to the contact area producing it.
(iii) Charging being due to some secondary cause, for example the presence of contamination.

The only experiments in which anything like 2×10^4 e.s.u./cm^2 has been obtained, are the experiments on the contact charging of quartz already mentioned, in which a charge density at the ‘ point ’ of contact of at least $8{\cdot}7\times10^3$ e.s.u./cm^2 was estimated. This charging must have been primary in character; the question of whether electrons or ions were transferred remains to be considered.

It must next be explained why the charging of insulators cannot be described by a simple application of the energy level diagrams shown in fig. 2, along the lines that account for the establishment of the contact potential between metals. Let us first consider an experiment of Perucca. He found that the contact potential of mercury against a standard metal surface changed with time as the fresh mercury surface aged in air, and that this was correlated with a change in its ability to charge a glass surface by contact. The ageing led to a reversal of the sign of charging, as shown in fig. 3. Suppose, now, the charging were due to a transfer of electrons between metal and insulator and that these can be represented by the level systems of fig 2, and also that the ageing process causes a steady shift in the relative heights of the two systems. Negative charging of the insulator would mean that the Fermi level in the metal (the top of the filled set of levels) was higher than the bottom of the empty (conduction) band in the insulator, electrons being able to fall into lower levels by passing from left to right. Positive charging of the insulator would mean that the top of the filled (valency) band in the insulator was higher than the Fermi level in the metal, electrons being able to fall into lower levels by passing from right to left. The progressive shift from negative to positive charging, however, takes the two energy level systems through the situation shown in fig. 2, in which the Fermi level of the metal is opposite the forbidden band of the insulator, and electrons cannot pass in either direction, there being no lower unoccupied level

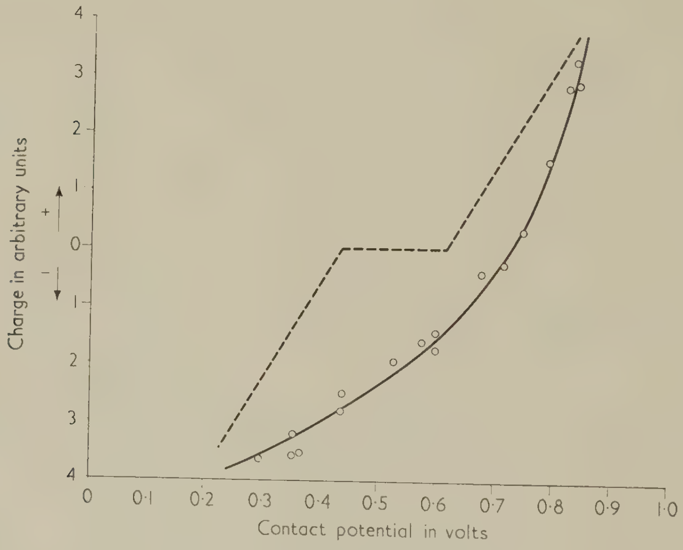


Fig. 3. Perucca’s experiment.

to which they can go. This would imply the existence of a range of contact potential of the mercury within which there should be no charging, and hence a charging graph similar to the dotted line in fig. 3. The observations actually fit a *smooth* curve, and are therefore incompatible with the simple hypothesis we have been considering.

Some results of Harper also call for a rejection of the simple hypothesis, for another reason: an example is shown in fig. 4, which refers to metal/insulator contact charging, the cumulative charge after a succession of contacts being plotted against the number of contacts. It is seen that gold charges more strongly than steel against magnesium oxide, the metals acquiring a negative charge. This would imply that the relevant electron levels were in the order shown on the left of fig. 5. The metals acquire a positive charge against glass which implies an empty level in the glass as on the right of fig. 5, but this order of levels would make steel charge more strongly (against glass) than gold: the experimental results are the opposite way round. It is clear that the experimental results of fig. 4 cannot be reconciled with a level system.

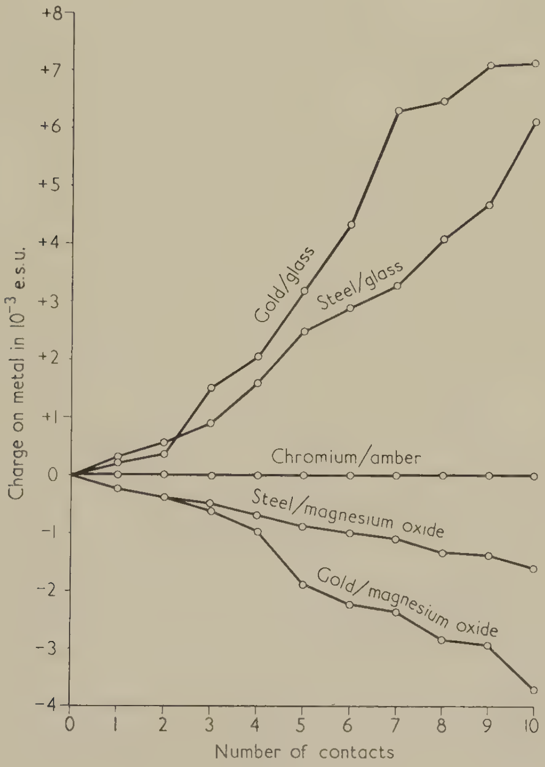


Fig. 4. Examples of metal/insulator contact charging.

It is also apparent that the materials of fig. 4 do not arrange themselves in a triboelectric series in which pairs which are further apart in the series charge each other more strongly than pairs which are close together. Many attempts have been made to arrange materials in such a series, but with varying degrees of success, which is not surprising when it is mentioned that even the sign of charging can be changed by using a different method of ‘ cleaning ’ the materials tested. Some regularities which have encouraged those who have looked for a triboelectric series certainly exist, but the regularities are provocative rather than helpful.

We must now ask how the energy level diagram of fig. 2 *should* be interpreted for contact charging. Thermodynamical considerations require that when two materials in contact have attained equilibrium, their Fermi levels are at the same height. The Fermi level is such that lower levels are more likely to be occupied than empty, and higher levels more likely to be empty than occupied. The Fermi level in a metal is almost exactly half way up the band, with effectively all levels below it occupied, and effectively all levels above it unoccupied. In an insulator, the Fermi level proves to be a mathematical fiction, in the sense that it is almost exactly in the middle of the *forbidden* band. An insulator/metal contact cannot be in thermodynamical equilibrium until the Fermi levels in the insulator and the metal are at the same height. If this adjustment is to occur, a double layer must be formed. The part of the double layer that is in the insulator will be a space charge made by a change in the number of charge carriers in the insulator. But the number of charge carriers in the conduction band of an insulator is negligibly small, whence the thickness of the double layer must be extremely large. Furthermore, owing to the negligible conductivity of the insulator, it would take an extremely long time for the double layer to build up. Calculation shows that the formation of such a double layer in ordinary laboratory experiments with insulators requires both a prohibitively long time and a prohibitive thickness of material, so that thermodynamic equilibrium is never reached. Expressed simply, since the double layer cannot form, the insulator will not charge. It is not in equilibrium, though it appears to be.

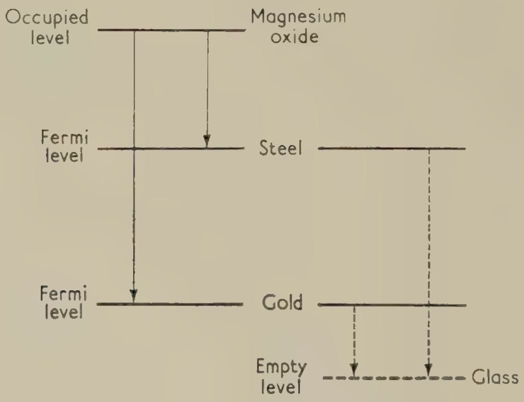


Fig. 5. Failure of electron level system to explain fig. 4.

We now understand that it must be the electrophobic insulator which behaves as a simple insulator ought: an insulator cannot conduct, how then should it acquire a charge? Experimental results for amber/chromium are shown in fig. 4. The charging is negligible; this was equally true for amber/gold, which excludes the possibility that, by chance, amber and the metals happened to be levelled up before contact—the work functions of the gold and chromium used in the experiments differed by 0·9 v.

Quartz and amber are both extremely good insulators, yet the one charges sufficiently strongly for the cause to be obviously primary in character, the other not at all. The theoretical argument that an insulator cannot charge by electron transfer applies equally to amber and to quartz. This consideration forces the conclusion that the charging of quartz is due to the movement of positive or negative *ions*, but that loosely bound ions are not present on the surface of clean electrophobic insulators. In general, surfaces may be able to transfer both electrons and ions, but it is the electrons that will ' have the last word ': the redistribution of ions stops as soon as the surfaces have become separated by a gap of molecular dimensions, whereas the redistribution of electrons can proceed by quantum-mechanical tunneling until the gap is larger by an order of magnitude. In such a case we would say that the primary cause of charging is the transfer of electrons, but that this is complicated by a change in the adsorbed surface layer following contact, because, if electrons can be transferred, positive ions that move will take electrons with them if this is required for stability, and the resultant charging must be accounted for by considering the altered electron levels. It is only when *electrons cannot be transferred* that the charging may properly be described by the movement of *ions* from one surface to the other. When the charging *is due to electrons moving*, the transfer will be between interior levels, since surface levels are almost certainly either fully occupied, or quite empty, shifting the interior levels up or down. Adsorbed positive ions with which electrons become associated will form a double layer at the surface, and do likewise. An example is the charging of metals on which there is an invisible layer of oxide. The constitution proposed for this layer by Cabrera and Mott leads to an increase in the work functions of the metals, though with smaller differences between them, so that contact potentials are less than for clean metals, as is observed. In confirmation of what has just been said, Harper's experiments on the charging of metals give quantitative agreement with the theory for clean metals *provided the comparison is made using the observed contact potentials for the metals with a surface layer.*

4. SEMICONDUCTORS

The band structure of an intrinsic semiconductor is like that of an insulator, but the forbidden band is narrower so that an appreciable number of valency electrons are raised into the conduction band by thermal agitation. This makes it possible for the Fermi level of the intrinsic semiconductor to be displaced with respect to another material with which it is in contact, by the formation of a double layer of charge, so establishing the contact potential. Owing to the limited number of mobile electrons the space charge cannot be concentrated near the surface: the lower the conductivity, the thicker the space charge will be (other things being equal). In the limit, as we have already seen, the space charge needs to be so thick that there will not be room for it, and in principle the semiconductor has become an electrophobic insulator. At the other extreme of conductivity, the space charge can be thin, and the semiconductor will behave very like a metal. Excess and deficit semiconductors are somewhat more complicated, but similar considerations apply, and with all types electrons can move and ' have the last word '. We therefore attribute the charging of semiconductors to the movement of electrons; if positive ions move, we say the nature of the surface has changed. Comparing the contact charging of a semiconductor with that of metals, the essential difference is that the component of the double layer in the semiconductor extends well below the surface. The greater thickness of the double layer means that a given potential discontinuity is provided by a smaller field within the layer, and therefore by a smaller charge. It follows that (other things being equal) the lower the conductivity, the less a semiconductor will charge. In principle, semiconductors fill the gap between metals and electrophobic insulators.

Not many experiments have been performed which permit a comparison with theory for semiconductors, but an explanation can now be given of an empirical law found in Professor M. B. Donald's laboratory at University College, for the case of rutile powder sliding down a metal chute. It was found that, for varying inclinations and length of chute, the charge on the powder was a function of the time of descent only. Now when an uncharged rutile particle comes into contact with a metal chute, a current will flow through the contact ' point ' because of the contact potential. The current will only stop flowing when the whole particle has reached the same potential, that is when the tiny capacitor formed by particle and chute has become charged to a potential equal to the contact potential. In the meantime, the charging current is severely limited by the constriction at the ' point ' of contact. This charging model, therefore, leads to the expectation that the powder will charge like a capacitor to which a potential is suddenly applied through a resistance: just what was found in the experiments on rutile. Rough quantitative comparisons support the explanation.

5. ELECTROPHILIC INSULATORS

Surface charging found in practice on insulators almost invariably falls between the two extremes which are obtainable in the laboratory by carefully cleaning the two types of insulator. That this is so is attributable to the effect of the two corresponding types of contamination. Electrophilic contamination on an electrophobic insulator will lead to some charging, whereas electrophobic contamination on an electrophilic insulator will obscure potentially high charging such as is found with clean quartz. In practice, then, both types of insulator show intermediate charging; if contamination is removed, they separate into the two extremes, electrophilic insulators charging more, and electrophobic less. It could be argued that to account for the moderate charging of a typical insulator such as mica by saying that it doesn't charge as much as an electrophilic insulator ought because it is usually contaminated by something electrophobic, is inventing a difficulty in order to dispose of it by another invention: what is the electrophobic contamination, and why is it so prevalent? Admittedly, it *is* a little difficult to account for, but grease will act in this way, and a great deal of atmospheric pollution is of a greasy nature. The author, however, suspects that the ' veils ' discovered by Hast may be the real answer. Hast found that crystals, even when one would expect them to be clean, seem often to be encased in an extremely tenuous skin, and that this skin can be preserved relatively intact when a special technique is employed for dissolving out the crystal, enabling the skin to be photographed in the electron microscope. An example of such a veil, from sodium chloride, is shown in fig. 6. Hast found that these veils, though difficult to detect, were in fact very prevalent. They have a fibrous, mat-like structure, and could well be electrophobic.

There is no need to look far for electrophilic contamination: water, particularly if containing a trace of electrolyte, is the obvious candidate. In this connection, an experiment performed by Deaglio in 1928 is significant. He investigated the potential existing between two metal balls, when the gap between them was varied in the range 0 to 1000 Å, in air of variable relative humidity.

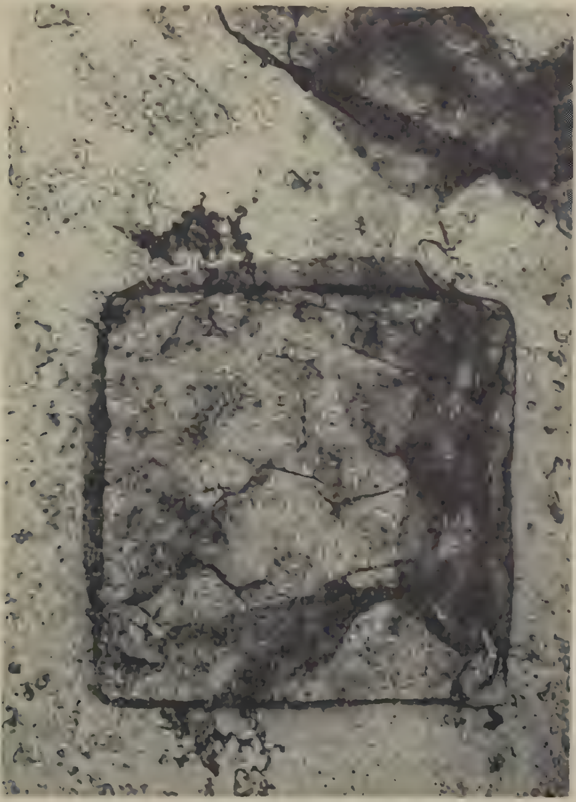


Fig. 6. Veil from sodium chloride.†

For sufficiently small gaps, of the order of 100 Å, and not too low relative humidity, i.e. in excess of the order of 5 per cent, the contact potential was replaced by a potential of opposite sign, and rather smaller in magnitude. This is what would be expected if the small gap became filled with electrolyte, so forming a battery with the two spheres as electrodes. Presumably there was a precipitation of water on a trace of hydrophilic contamination, thus causing the phenomenon, and there seems to be no reason why the contact charging of insulators should not sometimes involve electrolytic effects of a similar nature. Indeed, experiments on the charging of zircon powder by sliding down a chute, which gave the same charging law as for rutile, may be an example of it. The electrolytic contamination would provide both the e.m.f. required to drive the charging current, and the surface conductivity without which zircon, with its high volume resistivity, could not become charged as a capacitor. It should be noted that this model of the charging process visualizes the charge transfer as being caused by a potential difference, the movement of charge being *in the direction of falling potential.*

---

† I am indebted to Dr. N. Hast for this photograph.

The charging of electrophilic insulators is certainly not always dependent, however, on the presence of water, and, perhaps, seldom is. Coehn and Lotz, in 1921, carried out experiments in which refractory powders did not lose their ability to charge after having been baked out at red heat in a high vacuum. The treatment was quite adequate to guarantee the elimination of water. In conditions such as these, we must look for a simple transfer of ions to give the charging, brought about by diffusion or some other cause, *but not driven by an e.m.f.* Such a movement of charge gives rise to a potential difference *which at first opposes the transfer and finally terminates it.* The movement is caused by forces that are not electrical in origin: we may call it mechanical transfer.

The difference between charging by mechanical transfer, and charging due to a potential difference, is apparent when we compare the empirical law found for powders sliding down a chute with the results of Wagner from Professor Loeb's laboratory at the University of California using apparatus designed with an insulating sphere rolling inside a rotating metal cylinder. In the experiments with powders, the fact that the charge on the powder depended only on the time of descent implies that the charging rate was *independent of the speed* of the particles in the powder, and this is compatible with the charging current being due to a potential difference. In Wagner's experiments, the initial charging current was *proportional to the speed* of rolling; this was interpreted as meaning that charge was transferred instantaneously when contact was made and left behind when contact was broken, so that the current was proportional to the rate at which new contact was being made by the rolling, and therefore initially to the speed.

The difficulties in the way of testing working hypotheses by experiment in the realm of contact electrification are brought out in full measure by Harper's discovery that different crystal faces of the same material charge each other, but that the full extent of the effect only appears after scrupulous cleaning. Figure 7 shows results for quartz. It was established that the charging was not piezoelectric in origin; the actual cause has not been demonstrated, but is thought to originate as follows. Deep electron levels on a newly-formed quartz crystal surface, probably associated with oxygen in some form, are at once filled by electrons, so charging the surface negatively. The different chemical structures exposed by different crystal faces give rise to different numbers of these electron levels. Later, when the negative charges are neutralized by positive ions, such as hydrated hydrogen ions, derived from the surroundings, these loosely bound ions are present in different numbers on different crystal faces. If two different faces make contact, a redistribution can occur because of diffusion, or because the attraction to the two surfaces differs because of different dielectric constants giving rise to different image forces. Quartz is anisotropic, the dielectric constant perpendicular to different crystal faces being different, but the difference is in the wrong direction to account for the observed charging, so diffusion would seem to be the likely cause in this case. Probably, however, electrical image forces are the dominant factor in other cases. This would lead to the surface of higher dielectric constant becoming positively charged with respect to the other; an empirical generalization that was enunciated by Coehn in 1898. This so-called law is often broken by nature; when it is observed, we may suppose that is because the charging is governed by loosely bound positive ions controlled by image forces. Then, the insulators concerned will arrange themselves in a triboelectric series.

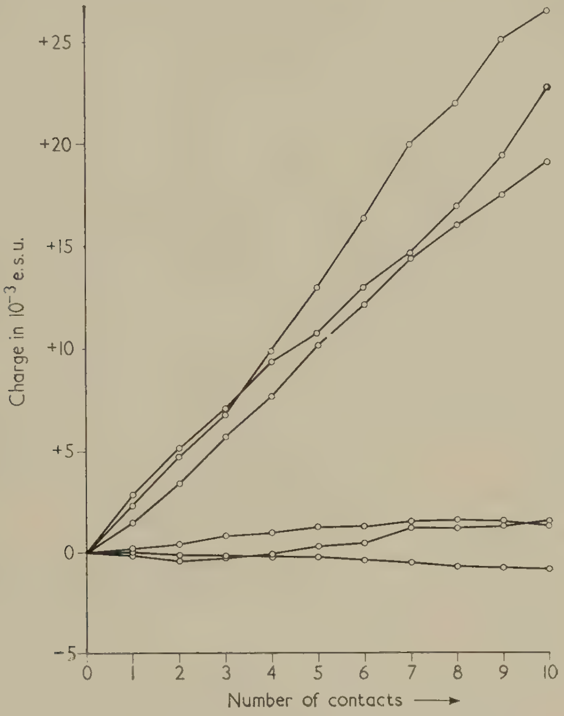


Fig. 7. Contact charging of quartz/quartz.
Upper plots: different crystal faces.
Lower plots: like crystal faces.

References

Arthur, D. F., 1955, A review of static electrification, *J. Text. Inst.*, **46,** T 721.
Harper, W. R., 1957, The generation of static charge, *Phil. Mag. Supp.*, **6,** 365.
Harper, W. R., 1960, Contact electrification of semiconductors, *Brit. J. appl. Phys.*, **11,** 324.
Henry, P. S. H., 1957, Generation of static on solid insulators, *J. Text. Inst.*, **48,** P5.
Loeb, L. B., 1958, *Static Electrification* (Berlin : Springer-Verlag).
Montgomery, D. J., 1959, Static electrification of solids, *Advances in Solid State Physics*, **9,** 139 (New York: Academic Press Inc.).

The Author :

Dr. W. R. Harper began research in the Cavendish Laboratory under the late Lord Rutherford and Sir James Chadwick; subsequently, in Bristol, stimulated by Professor A. M. Tyndall, he became interested in the problems of static electrification, but this was interrupted when he joined the Admiralty in 1939. During the second part of the war he was responsible to the Director of Naval Construction for the introduction of radiographic inspection of welding in ships under construction in British shipyards. In 1947 he again took up work on static electrification for the Ministry of Supply in Sir George Thomson's laboratory at Imperial College, where the experimental work which forms the basis of this article was carried out. He has recently written on nuclear reactors.

The Gyromagnetic Ratio of the Proton

by P. VIGOUREUX

The National Physical Laboratory, Teddington, Middlesex

Certain nuclei can for many purposes be thought of as spinning round an axis like the Earth or like a top. In general the spin endows them with angular momentum and with a magnetic moment ; the first because of their mass, the second because all or part of their electric charge may be rotating with the mass. Even the neutron, which does not have a net charge, has an electro-magnetic moment. (The electro-magnetic moment is a vector quantity, the vector product of which with the magnetic flux density is equal to the torque). The quotient of the electro-magnetic moment to the angular momentum, called the gyromagnetic ratio, would be calculable if to each element of mass were associated a corresponding element of charge, but as the value thus obtained turns out to be very different from the measured value, roughly 5·6 times smaller in the case of the proton, we conclude that our knowledge of nuclei is not yet adequate for the calculation of the ratio. As however the nuclear gyromagnetic ratio γ, especially that of the simplest nucleus, the proton, enters into many relations between atomic constants, its determination is a matter of some importance.

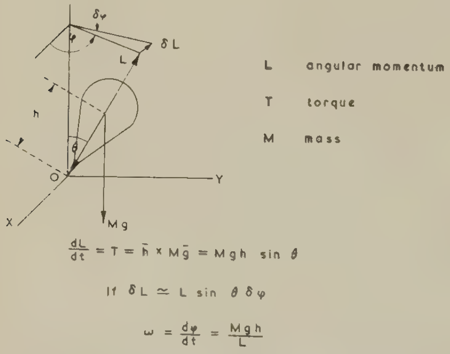

δφ
δL
L
T
h
θ
O
Y
X
Mg
L angular momentum
T torque
M mass
$\frac{dL}{dt} = T = \bar{h} \times M\bar{g} = Mgh \sin\theta$
If $\delta L \simeq L \sin\theta\,\delta\varphi$
$\omega = \frac{d\varphi}{dt} = \frac{Mgh}{L}$

Fig. 1. Slow precession of spinning top in gravitational field.

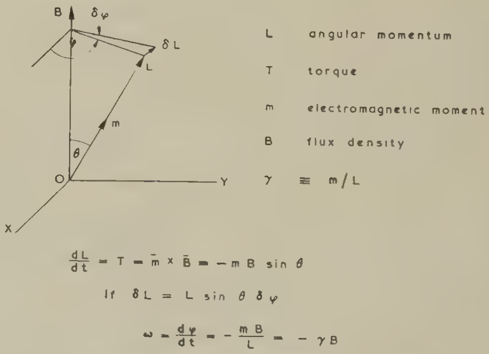

B
δφ
δL
L
m
θ
O
Y
X
L angular momentum
T torque
m electromagnetic moment
B flux density
$\gamma \equiv m/L$
$\frac{dL}{dt} = T = \bar{m} \times \bar{B} = -mB \sin\theta$
If $\delta L = L \sin\theta\,\delta\varphi$
$\omega = \frac{d\varphi}{dt} = -\frac{mB}{L} = -\gamma B$

Fig. 2. Precession of proton in magnetic field.

(*Crown copyright reserved.*)

The direct measurement of electromagnetic moment would not be easy, but the behaviour of the proton in a magnetic field leads to a simple method of determining the gyromagnetic ratio. A particle with electro-magnetic moment **m** in a region of flux density **B** suffers a torque **mB** sin θ tending to bring to zero the angle θ between the directions of the field and of the axis of spin. The effect is analogous to that observed for a spinning top with its axis inclined to the vertical : in this case gravity tends to reduce to zero the angle between the vertical and the axis, but because of the spin and consequent angular momentum, the polar angle θ, fig. 1, remains constant unless the top is jerked or otherwise disturbed ; what changes is the azimuthal angle ϕ, and the axis of the top is observed to rotate or ‘ precess ’ round the vertical. In the same way the axis of the proton precesses round the direction of the field, and it can be shown by a fairly simple application of mechanics, fig. 2, that the angular frequency ω of precession is equal to $\gamma \mathbf{B}$. It thus appears that a method of determining γ is to observe the frequency of precession of protons in a known magnetic field.

One way of obtaining a known field is to produce it by a known current in a coil of known linear dimensions. By international agreement the unit of current, the ampere, is defined from the force between electric currents and can be determined by some form of current balance in terms of the metre, kilogramme and second.

For the production of the field, fig. 3, a battery maintains current in a variable resistance r followed by the field-forming coils and by a standard resistance R whose potential terminals are connected to a standard cell E and a mirror galvanometer G. If r is varied so as to keep the galvanometer spot always at 'zero', the current I is equal to E/R ; and if the same resistance and cell, or others known in terms of them, are also used with a current balance, the current I is known 'absolutely', i.e. in terms of the metre, kilogramme and second.

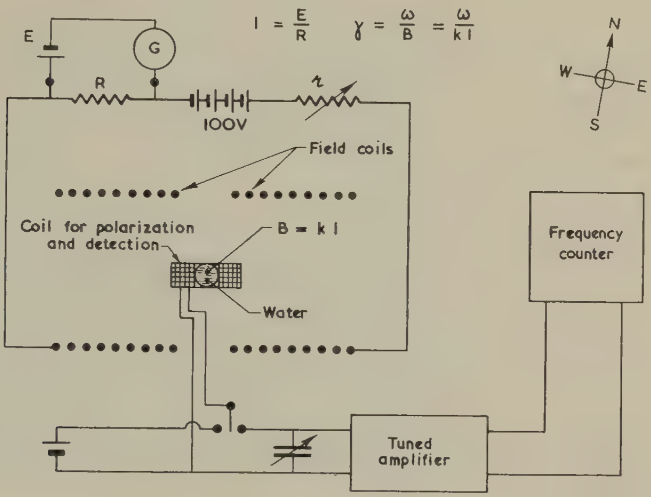


Fig. 3. Plan view of circuit for observation of free nuclear precession in weak field. (*Crown copyright reserved.*)

The next step is to calculate the flux density at the centre of the system, which can be done if the dimensions of the coil are known. To take a simple case as an illustration, if the coil were very long and uniformly wound with n turns, the flux density at the centre would be $4\pi \times 10^{-7}\, n\, I$ wb/m^2. In practice we cannot have an infinitely long coil, so the formula has to allow for the limited length, but it is then necessary to know not only the position but also the diameter of each turn of the coil. In one system a gap is provided in the middle in order to improve the uniformity of the field at the centre, on a principle analogous to that of the well-known Helmholtz arrangement.

A limitation imposed by the necessity of measuring the diameter and position of each turn of the coil is that only a single layer of bare wire is admissible, which means that the field is comparatively weak. Thus whereas in spectroscopy by nuclear magnetic resonance fields of the order of 1 wb/m^2 are common, our field is limited to about 2 mwb/m^2 with the result that the methods of absorption or of nuclear induction by which precession is normally observed in large fields are pretty well excluded here on account of the low signal/noise ratio obtainable in weak fields. On the other hand the method of free precession employed in the nuclear magnetometers so widely used of late turns out to be very convenient.

We have hitherto considered the behaviour of a single particle in a magnetic field ; but unless there are a large number the effect of precession on an external circuit is negligibly small, so it is usual to perform the experiments on a volume of the order of a cubic centimetre of substance. There are then billions of molecules each with thermal energies greatly exceeding the magnetic energy of the particles even in the strongest magnetic fields it is possible to produce. Thus, collisions due to thermal agitation soon produce a statistical equilibrium in which the average magnetization is biased, although only very slightly, in the direction of the field. The reason for the bias is that a particle has minimum potential energy when its magnetic axis is parallel to the field, maximum when it is anti-parallel. In water equilibrium is reached in a few seconds.

In the method of free precession a spherical cavity filled with a liquid rich in protons, e.g. water, is placed at the centre of the coil producing the standard field, fig. 3, and a strong field is applied in a perpendicular direction, for instance by current in a coil surrounding the cavity. All that happens at first is that precession occurs round this polarizing field, or to be exact round the resultant of it and the standard field ; but soon a new equilibrium is reached, with a bias in the direction of the resultant field—for 0·1 wb/m^2 at room temperature the bias is equivalent to only 1 proton in 3 million. If now the polarizing field is removed in a time short compared with the period of precession, this small magnetization at first keeps its inclination, i.e. about 90°, to the standard field, and precession occurs about the latter with a polar angle of nearly 90°, the most favourable condition for induction of e.m.f. in a detector coil at right angles to the field. The polarization coil is sometimes used for detection, as in fig. 3. But thermal energy, as explained above, works to restore the conditions proper to the weak standard field ; and the signal decreases exponentially, with a time constant which varies widely with the substance used, and for pure water at room temperature is of the order of 3 sec.

It might be thought that with excess protons in such a small proportion as 1 in 3 million the signal even at the start would be very small, but the signal can be calculated and turns out to be of the order of 1 μv induced in a coil of 1000 turns surrounding a water sphere 20 mm in diameter in a field of 1 mwb/m^2 after application of a polarizing field a hundred times as large, fig. 4. This signal, enhanced somewhat by tuning the detector coil, is taken to an amplifier followed by some device capable of measuring frequency to the required accuracy in a second or so. As in measurements of this sort the frequency to be expected in a given field is known beforehand to a fair accuracy, most of the noise in the system can be rejected by incorporation of a narrow-band filter in the amplifier—very simple circuits indeed will serve the purpose—and the measurement of frequency becomes easier and more accurate. Frequency counters are now available which operate by measuring the time of a suitably large number of periods of the input. If for instance the signal has a frequency of the order of 60 kc/s and decays in about 2 sec, the counter is made to measure the time occupied by say 80 000 periods. A counter is normally provided with a quartz crystal to serve as frequency standard, but if the measurements are made in a laboratory which maintains still better frequency standards, these can be used instead to operate the counter. It is moreover a simple matter also to produce from those standards a decaying curve closely resembling the precession signal in amplitude, frequency and time constant, and inject it into the detector coil to make doubly sure that the counter reads correctly. Errors in the measurement of the period of precession are thus almost completely eliminated, and indeed the accuracy of measurement of the gyromagnetic ratio is normally that to which the magnetic field is known in terms of the metre, kilogramme and second.

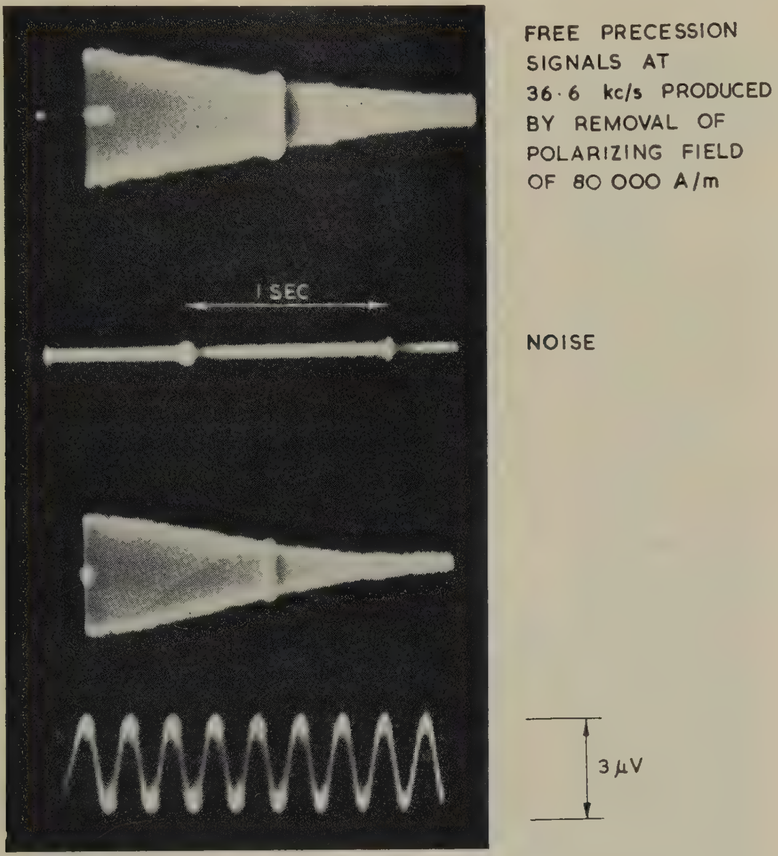


Fig. 4. Signals due to free precession of protons in water at 36·6 kc/s. (*Crown copyright reserved.*)

The frequency of precession depends not on the standard field alone but on the resultant of it and any other fields which happen to exist at the centre of the system. In a satisfactory site the Earth's field is by far the largest of these and its effect must be allowed for somehow. At the National Bureau of Standards(1), U.S.A., and at the National Physical Laboratory the axis of the field-forming coil is horizontal and perpendicular to the magnetic meridian, and two separate Helmholtz systems, fig. 5, are used to compensate the vertical and the horizontal components of the Earth's field. Even if the compensation were defective by as much as 1 per cent and the standard field were only 20 times the Earth's field, the small perpendicular residual field would cause an error of less than 2 in 10^7. There is of course the possibility that the axis of the standard field is not exactly at right angles to the magnetic meridian, but the small axial component of field thus introduced is eliminated by reversal of the current in the field-forming coils, a reversal necessary in any case to cancel also the stray fields due to the rest of the circuit.

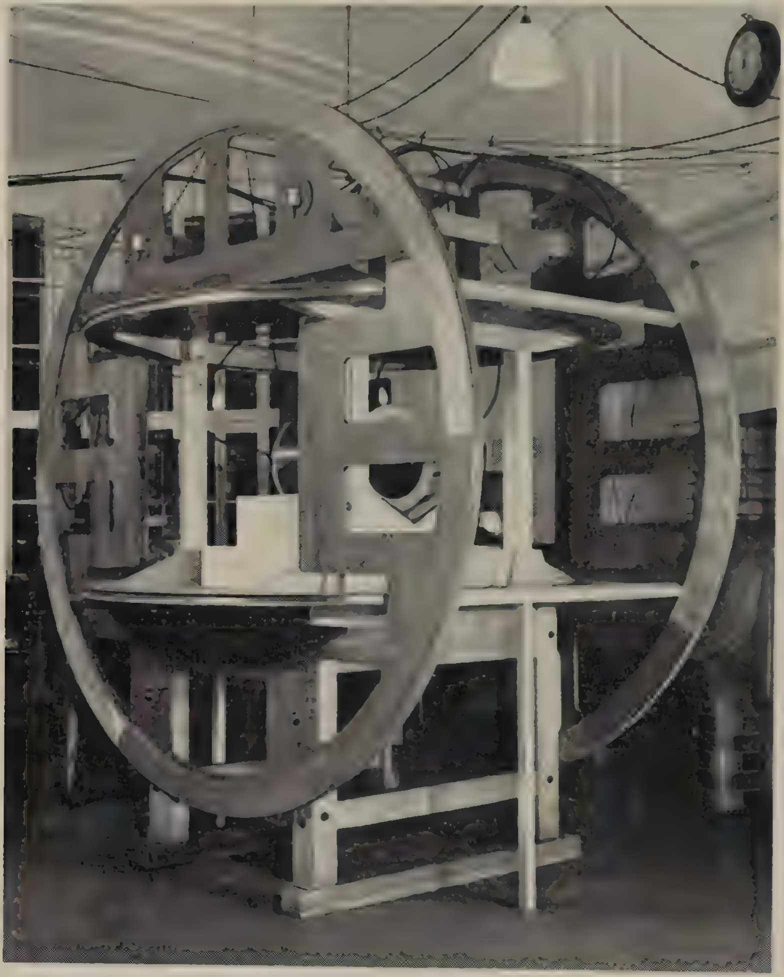


Fig. 5. Apparatus for measurement of the gyromagnetic ratio of the proton at the National Physical Laboratory. The four coils on the large wooden formers neutralize the Earth's field. At the centre is the field-forming coil on a marble former. (*Crown copyright reserved.*)

In another determination(2) for which the standard field was appreciably larger than for the measurements referred to above, compensating coils were dispensed with but the axis of the field-forming coils was set in line with the Earth's field, which again was eliminated by reversals.

But the Earth's field varies, and although the amplitude of the variation is small, the frequency spectrum is very broad. Moreover some industrial activities, e.g. electric traction, produce variable magnetic fields. These are avoided if work on nuclear magnetic resonance in weak fields is carried out away from industrial centres or at times of minimum activity, unfortunately as a rule in the small hours of the morning. Variations of terrestrial magnetism cannot easily be avoided, but they are not in general so severe as to spoil the measurements. Those with periods less than a tenth of a second do not affect a measurement of frequency taking 1 sec, whilst those with periods exceeding 10 or 20 sec can be allowed for by reversals of the standard field. Thus only a comparatively narrow band of the spectrum of variations is likely to be troublesome.

Another source of difficulty is the field gradient caused by magnetic material lying near the apparatus, for it gives rise to different frequencies in different parts of the substance, thereby shortening the signal by phase cancellation. Moreover in the calculation of flux density it is not in general possible to apply a correction for the effect of neighbouring masses of iron, which are therefore liable to introduce a systematic error in the determination. The obvious if inconvenient solution is to house the apparatus in a building free from ferromagnetic material.

One last point provides a good illustration of the methods of nuclear magnetic resonance. If the polarizing field is not reduced to zero in a time short compared with the period of precession in the standard field, the protons which it had aligned do not stay aligned, but follow instead the direction of the resultant of the decreasing polarizing field and the constant standard field ; thus they end aligned in the direction of the latter, and there is no precession signal.

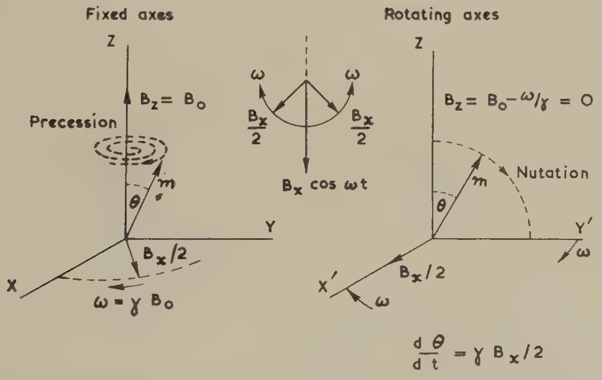


Fig. 6. Nutation caused by field rotating at precession frequency. (*Crown copyright reserved.*)

It is however possible to change the direction of magnetism, without affecting its magnitude, by applying a field at precession frequency at right angles to the standard field. This alternating field $\mathbf{B}_x \cos \omega t$ can be regarded as the resultant of two fields of half its amplitude rotating in opposite sense at precession frequency in the plane perpendicular to the standard field. Of those two fields the one rotating in the sense opposite to precession has but negligible effect on the average, the other, fig. 6, causes a nutation of extent proportional to the product of the rotating field and the time for which it is applied. The ideal is to set the magnetic axes of the protons at right angles to the standard field, let them precess and measure the frequency. This 90° rotation needs a field–time product π/γ in terms of the amplitude of the applied alternating field. At the National Physical Laboratory this impulse is provided by discharging a condenser in the detector coil. The initial peak flux density of the damped train has to be $\pi\mathbf{B}_0/2Q$ where $\mathbf{B}_0$ is the standard flux density and Q the quality of the oscillatory circuit. The condition is reached by trial-and-error adjustment of the charging voltage, but is not a bit critical.

Measurements on the lines described above are being made at the National Physical Laboratory but it will be a few months before results can be published. Meanwhile it is worth noting that the Association of Geomagnetism and Aeronomy has recommended(3) that, pending agreement on a final value by an appropriate international scientific organization, all measurements of the geomagnetic field with a proton free precession magnetometer, using pure water, should be based on the value $2{\cdot}675\,13 \times 10^8$ wb^{-1} m^2 s^{-1} for the gyromagnetic ratio.

Some of the work described above is in progress at the National Physical Laboratory, and this article is published by kind permission of the Director.

REFERENCES

(1) BENDER, P. L., and DRISCOLL, R. L., 1958. A free precession determination of the proton gyromagnetic ratio. *IRE Transactions*, **I**–7, 176.

(2) WILHELMY, W., 1957. Eine Neubestimmung des gyromagnetischen Verhältnisses des Protons. *Ann. Phys., Leipz.*, **19**, 329.

(3) *Nature, Lond.*, 1960. **188**, 192.

The Author :

Studied at City and Guilds College, South Kensington, has published books on Electric and Magnetic Measurements, on Quartz Oscillators, on Ultrasonics and papers on Acoustics, on Electrical Measurements and on Electric Units and Standards. Is now in charge of the Electrical Standards and Measurements Group (Standards Division), National Physical Laboratory.

The Scalpel and the Spectroscope

by J. M. A. LENIHAN
Western Regional Hospital Board, Glasgow

1. INTRODUCTION

The object of this paper is to comment on the close connection between the study of physics and the practice of medicine. The partnership is of respectable antiquity. Indeed, it could hardly be older, for it originated in the school of Aristotle. He regarded φυσικη as the study of natural phenomena, organic and inorganic; so the same term came to be used for science and for the healing art—as well as for metaphysics, which was the subject next to physics in the curriculum of the Lyceum at Athens where the Peripatetic philosophers held their classes.

The word physicist was coined in 1840 by William Whewell, professor of Moral Theology at Cambridge. It did not pass into the language without protest. A writer in Blackwood's Magazine complained in 1843 about "the word *physicists* where four sibilant consonants fizz like a squib". Guided perhaps by this admonition from Edinburgh, the Scots have avoided further objections by continuing to study Natural Philosophy.

2. PHYSICS AND PHYSIC

For two thousand years the art of healing and the advancement of science were practised by the same people. While science and medicine were both ruled by authority and tradition the same training and temperament would serve for either; this was true until well into the 17th century. Then, with the development of modern science, based on enquiry and analysis, the old structure of Aristotle's physics began to fall apart. A formal distinction between medicine and natural philosophy became common during the 17th century, and chemistry emerged as a separate science during the 18th century.

During the 17th and 18th centuries the relation between physics and medicine was still very evident. There were no professional physicists in those days. Until quite well into the 19th century important advances in physics were being made by men who had been trained as doctors; there was no other way to gain a knowledge of science. Systematic training in physics was an idea conceived by two Scottish mathematicians—William Thomson (later Lord Kelvin), who made the first students' laboratory in a Glasgow cellar in 1850, and James Clerk Maxwell, who encouraged the Cambridge wranglers to apply their mathematical learning to practical problems.

The present relationship between physics and medicine depends on three factors, of which two are fairly obvious and the third is more subtle. A backward glance will show them in perspective.

In the long succession of doctors who helped to found the science of physics the first was William Gilbert of Colchester (1540–1603), physician to Queen Elizabeth I and later to King James I and VI. Navigation—an art of national importance in those times—was the inspiration of Gilbert's work on the magnet.

Using a lodestone sphere and a small compass, he was able to demonstrate many effects known to sailors and to develop a model—little improved for centuries after—of the earth's magnetism. He also studied electrification (having coined the word *electric*) and made the first electroscope. His writings show a clear appreciation of experimental method and his great work *De Magnete* was the first book on physics written by an Englishman.

Gilbert's medical practice seems to have been conducted in orthodox fashion. For all his inside knowledge he did not dabble in the therapeutic applications of magnetism which enriched less reputable physicians even into the 19th century.

3. The sage of Darmstadt

If we except the dubious practices associated with magnetism, the first branch of physics to find a use in medicine was electricity. It is strange to recall that, for a century and a half after the early discoveries of Gilbert and his contemporaries, no one regarded electricity as anything but an amusing novelty. The idea of practical usefulness was quite late in appearing. It was due to Johann Gottlob Krüger, a little known scholar who was professor of philosophy and medicine first at Halle and later at Darmstadt. The remnants of his work now surviving show him as a wise and witty teacher. In the autumn of 1743 he was asked by his students to discuss the possible utilization of electricity.

" God only knows " (he said) " what the ingenious heads of our times will get out of it all . . . . If it must have some practical use, it is certain that none has been found for it in Theology or Jurisprudence, and therefore where else can the use be than in Medicine ? "

Krüger was not joking. He proposed a scheme of investigations on the biological effects of electricity, and his students were the first to do serious work in this field. The difficulties were formidable, for the only source of electricity was the friction machine—generally a glass globe or bottle, excited by the experimenter's hand. Krüger knew that electricity could flow through the body and supposed that various therapeutic effects must follow. His predictions of muscle contraction and increased pulse-rate were verified in 1744 by Kratzenstein of Halle who quickly became expert in the electrical treatment of stiff or palsied fingers using a technique which, with refinements, is still serviceable.

The new ideas were taken up in many countries. The foremost British investigator was another doctor—William Watson of London. He was the first to use the earth return in an electrical circuit and to formulate the idea of electrical resistance; in 1747 he enunciated a principle anticipating Ohm's Law. Watson's books, sent to Philadelphia in 1746, led Franklin to start his own experiments in electricity.

4. Measurement in medicine

Apart from the realization that science might be useful there was another revolutionary idea abroad in the 18th century and indeed in the 17th century—the idea of measurement in medicine. In recent years this idea has found many interesting applications. The progress of experimental physics during the last two decades has been accompanied by great improvements in the accuracy of measurement. Simple quantities such as times, displacements, velocities and frequencies as well as more subtle quantities such as radiation and ionization can now be estimated with an ease and precision that were not attainable in the 1930's. The new techniques have so many uses in physiology and biochemistry that measurement is a well-nigh universal constituent of clinical science today. It was not always so. The first man to introduce it into medical practice was Sanctorius of Padua (1561–1636). He was the first physician to measure the pulse-rate. As there were no portable timepieces, he devised a *pulsilogium*. This was simply a pendulum, adjusted in length until its beats were synchronous with the patient's pulse.

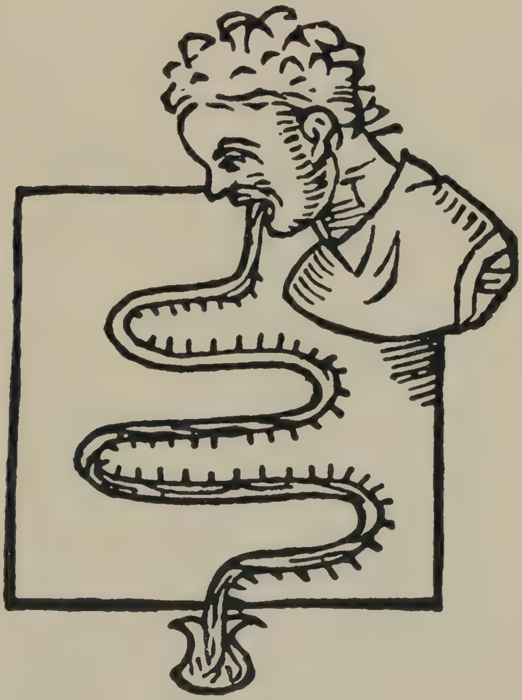


Fig. 1. The clinical thermometer used by Sanctorius of Padua (1561–1636). He was the first physician to appreciate that the temperature of the body might be a useful diagnostic sign.

Sanctorius was also the first to appreciate that the human body maintained a constant temperature in normal health and that deviations from it were of diagnostic value to the physician. The instrument that he used for estimating temperature consisted of a bulb (placed in the mouth or held in the hand) connected by a tube to a vessel containing water or wine. The fall of the liquid level in the tube during a standard interval (generally 10 pulse beats) was an indication of the body temperature.

The thermoscope of Sanctorius was similar in design to the air thermometer of Galileo and the priority of invention is sometimes disputed. Both instruments were, of course, much nearer to the ancient devices of Philo of Byzantium and Hero of Alexandria than to a modern thermometer. The significance of Sanctorius's work is in the use that he made of the instrument, rather than in its invention.

He was also a pioneer in metabolic studies, made with the famous steelyard on which he often ate and slept, sustained—as a familiar woodcut shows—by meat, wine and other needables. He weighed everything that passed into the body and everything that came out, thus estimating the loss by what he called " insensible perspiration ".

Further developments in the study of metabolic processes had to await the invention of an efficient calorimeter by Lavoiser and Laplace in 1780. Measurement of the pulse-rate did not become a regular part of the physician's examination until the 19th century. The modern type of clinical thermometer, with a constriction to prevent the immediate return of the mercury to the bulb, was patented by Negretti and Zambra in 1852 and the instrument familiar today was introduced into medical practice by Allbutt in 1870.

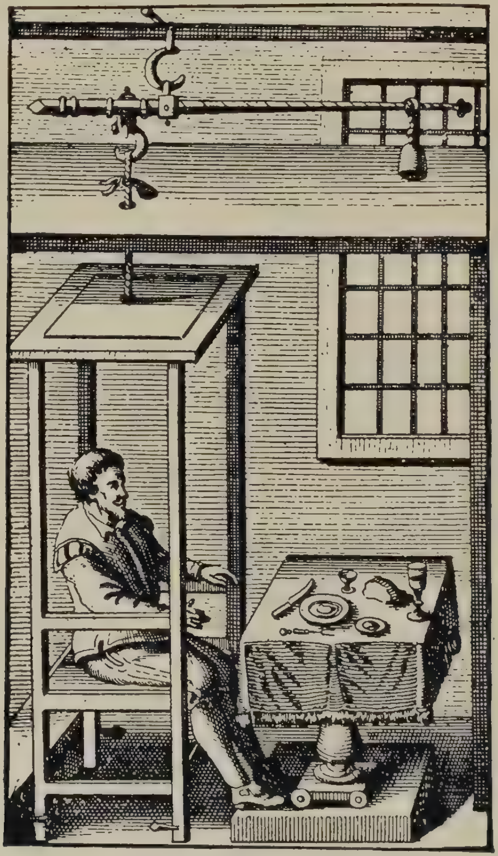


Fig. 2. Sanctorius seated on his steelyard, where he made his pioneer experiments on the metabolic processes of the body.

Sanctorius may be claimed as the father of medical physics. His successors use radioactive isotopes for more accurate metabolic tests. We use electronic circuits for measuring the pulse-rate and thermocouples for recording the body temperature but the basic idea behind these techniques—the idea that the physician can learn something useful by making measurements—we owe to Sanctorius.

5. KELVIN AND THE LINEAR ACCELERATOR

Let us now look at the works of a few men who have contributed richly to medicine and to physics during the last 400 years. We should perhaps start with Kelvin, who was a true disciple of Sanctorius with a profound respect for the importance of accurate measurement.

His contributions to clinical science were not obvious. He never studied medicine, though on one occasion at least he claimed a measure of competence in the healing art. The University of Heidelberg, at its tercentenary in 1884, wished to confer an honorary degree on him. Kelvin was already many times a doctor of laws and of science; the only further distinction that Heidelberg could give was the doctorate in medicine and this they offered him very willingly. In 1893, Kelvin was a witness in a lawsuit arising from fraudulent claims made by the manufacturers of an " electric belt ". The advocate who cross-examined him began by acknowledging Kelvin's eminent qualifications but observed sarcastically that none appeared relevant to a dispute on medical matters. " Well," said Kelvin, " I am a doctor of medicine of Heidelberg ! " His adversary was baffled and troubled him no more.

For our present purpose, Kelvin's most important work was his unknowing contribution to the development of the linear accelerator, an instrument which did not appear until long after his death. Now used extensively as an x-ray source for the treatment of cancer, it is a good example of a trend which has become increasingly evident during the last 30 years. During this time nearly every high-voltage electrical machine has been made in two versions—one for the nuclear physics laboratory and one for the hospital. Fortunately for the taxpayer, this process ceases to be attractive at energies above 50 MEV.

The linear accelerator has a curious history, linked with the Glasgow, Paisley and Ardrossan Canal. This waterway, begun in 1806, reached as far as Johnstone, near Paisley, when the funds ran out; a remnant survives in the grounds of J. and P. Coats's thread mill in Paisley. The canal was conducted with great enterprise until it succumbed to competition from the railway in 1881.

6. SAGACITY OF A HORSE

On a certain day in the 1830's a horse was towing a barge along the canal at the accustomed speed of about 4 m.p.h. Suddenly the horse took fright and bolted, pulling the boat behind him. Travelling faster and faster, to the alarm and anxiety of the bystanders, who expected him to collapse from exhaustion, the spirited beast attained the unprecedented speed of 12 m.p.h. and finished the journey at a steady gallop, arriving undamaged at the terminus. This splendid achievement was more than the proprietors could understand, but they quickly organized an express service of fly-boats drawn by pairs of horses. As soon as a speed of 12 m.p.h. was reached, little effort was needed for the horses to continue.

Sir William Thomson, who was keenly interested in navigation and ship propulsion, provided an explanation in 1886. At low speeds, a canal barge is preceded by a succession of waves. The power needed to produce these waves represents a major part of the work done in towing the vessel, and is, of course, provided by the muscular exertion of the horse. The waves have a characteristic speed which depends only on the depth of the canal. If the barge can be accelerated to this speed it makes no more waves but rides on a single crest with very small resistance.

Sixty years later a similar technique was used for the acceleration of electrons. A radio signal of short wavelength was projected into a hollow metal tube and reduced to a relatively low speed by a series of diaphragms restricting the opening of the tube. A supply of electrons from a heated wire was injected at the same time and place, having first been accelerated to the speed of the wave. The electrons rode on the wave which was then allowed to attain its full speed. The electrons riding with it were soon travelling at near the speed of light, with very high energies.

This was the prototype of many machines which have since been built using the same principle. The linear accelerator is probably the only scientific instrument to have been invented by a horse. Let us now consider some of the contributions made to medicine by British physicists of the human kind.

7. A BREATH OF LIFE

Our next topic—the physics of breathing—has a history which goes back beyond Sanctorius and even beyond Aristotle. The act of respiration is the most obvious external sign of life and it is not surprising that methods of resuscitation have long exercised the attention of both doctors and scientists. So far the various artificial means of respiration are far behind the natural method in simplicity and reliability. Indeed, it is only during the last few years that any substantial technical effort has been devoted to this problem.

Artificial respiration is one of the oldest clinical techniques. In 2 Kings, 4, 34, we learn how Elisha revived the child of the Shunammite woman using the method of intermittent positive-pressure ventilation:

" And he went up, and lay upon the child, and put his mouth upon his mouth, and his eyes upon his eyes, and his hands upon his hands, and he stretched himself upon the child; and the flesh of the child waxed warm . . . and the child sneezed seven times, and the child opened his eyes ". It must be admitted that some Old Testament scholars dispute the technical virtuosity of Elisha's achievement, explaining it as merely an act of mimetic magic.

The next advance was made by Paracelsus, who used his fireside bellows, heedless of the hazard from inhalation of cinders. After that the positive-pressure method was revived about a century ago and has since become familiar to anaesthetists. It was first used for the routine treatment of patients in the Danish poliomyelitis epidemic of 1952–3. More than 3000 people were affected and respirators of the iron-lung type were being used on a large scale with little benefit. The mortality rate was about 90 per cent even in patients supplied with ample quantities of pure oxygen. The trouble was that carbon dioxide was not being eliminated from the blood.

At this crisis, Professor H. A. C. Lassen decided to abandon the respirators. Each patient was fitted with an endotracheal tube through which a mixture of oxygen and nitrogen was delivered under positive pressure. The motive power was provided by teams of medical students squeezing rubber bags at an appropriate rate to direct gas from storage cylinders into the patients' lungs. By this means the course of the epidemic was changed and the mortality rate among patients with respiratory paralysis dropped to 30 per cent.

Forced ventilation of the lungs at regular intervals, as applied by the simpler types of mechanical respirator, is most unwelcome to a small baby, who always breathes in an irregular way. Experiment shows that effective ventilation of the lungs can be achieved only when the incoming draught of air arrives at the same time as the child's own effort at inspiration. It is therefore necessary to make a servo respirator, in which a signal generated by the patient is used to set off a train of events which result in the delivery of a suitable volume of air at a pressure sufficient to inflate the lungs. The Glasgow machine developed since 1955 uses a mask to detect the small pressure change when the child tries to breathe. This pressure lifts a sensitive flap forming part of a photo-electric trigger circuit delivering an electrical impulse to the electronic timing circuit. The timer controls the outlet from a cylinder of gas through a solenoid valve. In this way the infant can indicate when he wants a dose of air, but the amount delivered is not limited by his own feeble strength. The servo respirator, which has had a great success in the treatment of respiratory distress in the newborn, forms the basis also of an automatic anaesthetist now being developed in Glasgow.

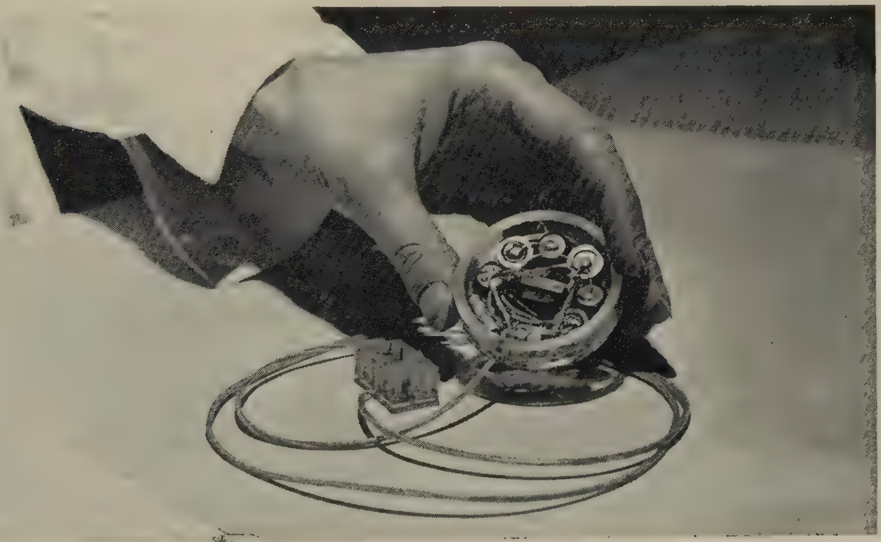


Fig. 3. Prototype of an electronic pacemaker. The electrodes, which are attached to the heart, can be seen projecting from the rectangular capsule. (*Courtesy of The Glasgow Herald*.)

During the development of the servo respirator, instruments were contrived for the measurement of various quantities involved in the breathing process. One of the more exciting experiments made in this work was the recording of pressures generated during the first breath of a baby's life. A normal adult needs to generate a negative pressure of about 3 cm of water to keep his lungs ventilated. A newborn baby produces a pressure of 60–80 cm of water in his first breath, an achievement not likely to be equalled during the rest of his life—unless he grows up to be a trumpeter.

An electronic implantable pacemaker, which will keep the heart beating, has been designed and built at the Regional Physics Department. Fixed inside the body, with its electrodes attached to the heart inside, it delivers timing pulses (from a transistor oscillator) of 1–2 volts amplitude, 1 millisecond duration and about 5 mA peak current. The pacemaker is used in cases where the normal timing action of the heart has failed. The model shown in fig. 3 runs for a year without attention, but new types now under development will need attention for battery replacement only every three years.

8. Hooke and Black

The next figure in our story, Robert Hooke (1635–1703), showed the breadth of vision and achievement that were common among the men who practised physics and physic in earlier times. His scientific work, mostly done as Curator of Experiments to the Royal Society, makes an exciting catalogue. Apart from the well-known work on elasticity, he invented the universal joint and, by the introduction of the anchor escapement and the balance spring, made possible the subsequent construction of accurate clocks.

As one of the Surveyors of the City of London he directed the rebuilding after the Great Fire of 1666. He helped Wren with many of the city churches, including St. Paul's and, but for his other interests, would certainly have been known to us as a famous architect.

His medical practice contained little of science and a great deal of the superstition that was then prevalent. Mrs. Tillotson told him that " A soveraine remedy for the falling sickness was made out of the mosse of a man's scull" and down it went in his diary. Though no more skilled in diagnosis and treatment than any of his contemporaries, he yet made handsome contributions to medical science.

After improving the microscope by adding the field lens, he used it to study the structure of plant tissues. Here he discovered cells—so named because he thought that they resembled monks' apartments in a monastery. He was the first to see bacteria under the microscope—an achievement often attributed to Leeuwenhoek. As a chemist, he was the first to establish the rôle of oxygen in combustion.

Combustion was one of the interests of Joseph Black, who lived from 1728 to 1799. He is known as a chemist who discovered latent heat and specific heat; yet by training, inclination and experience he was primarily a physician. He helped to found the Glasgow school of medicine and later became an outstanding figure in Edinburgh. He was also in a very real sense one of the originators of the Industrial Revolution which, if it started anywhere, started in the Old College of Glasgow between 1757 and 1759. During those years Joseph Black was revealing and explaining the properties of steam, James Watt was thinking about the practical applications of these discoveries and Adam Smith, author of *The Wealth of Nations*, was preparing the economic foundations of an industrial society.

9. Thomas Young

In Black's last year as a teacher in Edinburgh one of his students was Thomas Young, another man of extraordinary versatility. He was already a fellow of the Royal Society of London when he arrived in Edinburgh in 1794 and matriculated in the faculty of medicine. After a year in Edinburgh he moved on to Göttingen and then to Cambridge, returning in 1799 to London where he took a house in Welbeck Street and put up his plate. In 1811 he was elected physician to St. George's Hospital, an appointment which he held to the end of his life. Young is best known for the wave theory of light which he put forward in 1801 but he did much more. He discovered the astigmatism of the eye and made the first successful theory of colour vision. He invented the ripple tank, used in every physics lecture theatre to show the properties of waves to students, and the kymograph, a familiar object in every physiology laboratory. He was the first to decipher the demotic and hieroglyphic inscription on the Rosetta stone, an achievement on which the modern study of Egyptology was based. He wrote a substantial part of the *Encyclopaedia Britannica* in its 4th edition, which was published in Edinburgh and, towards the end of his life—which was not a very long one—he became a scientific adviser to the Admiralty, directing the affairs of the Nautical Almanac Office and the Board of Longitude.

The *Nautical Almanac* produced under Young's direction was not renowned for its scientific accuracy, though it had other uses. After the end of the Peninsular War in 1814, Britain and Spain enjoyed a period of anxious peace, rather like the cold war of recent years in Europe. On one occasion (according to an unfriendly pamphlet by Sir James South, published in 1829) a ship of the British Navy, making a survey of the Mediterranean, received a courtesy call from a Spanish captain. He presented the commander of the British vessel, Captain W. H. Smyth, with a handsome silver dish. Captain Smyth, having no suitable souvenirs at hand, returned the courtesy by presenting his visitor with two volumes of the *Nautical Almanac*. This was an unreliable aid to navigation, for it was full of errors and even omitted 29 February in one leap year. No Englishman ever dared to use it but the Spaniard took it gratefully, sailed away and was never heard of again. Captain Smyth, using French and Italian navigational tables, returned safely to England and lived to a ripe old age as an Admiral.

After Young, the picture changes. During the 19th century there were a few men—Helmholtz is the most noteworthy—who achieved eminence both in medicine and in physics, but the continuing growth of knowledge has made such versatility unattainable. The doctor can now hardly ever make a direct contribution to the advancement of physics and for a time the physicist's contribution to medicine was much reduced.

More recently—about 30 years ago in England—a new collaboration began to grow around the therapeutic uses of x-rays and radium. The great problem here was the old one of measurement. In the 1920's there was no satisfactory unit of ionizing radiation and consequently no accurate means of prescribing treatments. The help of the physicist was sought in overcoming this difficulty. Before long the röntgen unit was established and a range of measuring techniques had been devised. The physicist now has a very important rôle in the planning and control of treatments given in radiotherapy departments. As a natural extension of these duties, he is able to supervise radiation protection and to give technical help in the use of radioactive isotopes. Here we may make a brief historical digression.

10. Radioactive isotopes

The emperor Vespasian (who reigned from A.D. 70 to 79) holds an important place in the history of medicine. He made great improvements in the Roman health service and was one of the first to pay doctors' salaries out of public funds. The cost of these innovations did not worry him much for he displayed a bizarre fiscal ingenuity. His most notable achievement was the lavish provision of outdoor toilet facilities with storage tanks for the urine, which he sold to the dyers and launderers, thereby advancing the level of public sanitation while enriching the imperial coffers. This notable financial enterprise is commemorated today by the vespasiennes* which stand in the streets of Paris for the public convenience. Vespasian's son, afterwards the emperor Titus, chided him for raising money by such unpleasant methods; in reply he held up a bag of gold and uttered the celebrated Latin tag—*Non redolet.* Vespasian would surely have imposed a swingeing tax on those hospital laboratories where large amounts of bodily refuse are daily examined for the benefit, though not the profit, of patients and their physicians, for the discarded fluid is valuable as a means of studying the metabolism of radioactive materials administered to patients.

Many elements can be changed into radioactive varieties, or isotopes, by exposure to neutrons inside a nuclear reactor. The radioactive isotopes have the same chemical and biochemical properties as the stable parent elements but have the additional facility of emitting radiations which can be detected by Geiger counters or similar devices. Radioactive assay in general is a simpler and quicker process than chemical analysis. Consequently many interesting and important investigations can be made on topics beyond the reach of conventional methods. An outstanding example is the use of radioactive iodine in the study of thyroid disorders. The thyroid gland contains 5 to 10 mg of iodine, a substantial part of the body's store. Changes in thyroid function are reflected in the turnover rate of new iodine administered to the patient. If a small amount of radioactive iodine is given by mouth, its progress can be studied by radioactive assay of urine, of blood (for protein bound iodine, i.e. thyroid hormones) or of the intact thyroid gland, using external counting equipment.

11. Activation analysis

Another useful application of radioactivity is in activation analysis. Most of the chemical and physical properties of matter depend on the electrons which revolve round the nucleus. They are responsible for chemical combination, for mechanical strength, for conduction of heat and electricity, and, in more subtle ways, for virtually every observable property of stable matter. The nucleus takes no part in any of the ordinary processes of chemistry of physics. It can however be made to demonstrate characteristic properties by exposing a sample of material to neutron bombardment, thereby making a proportion of the nuclei into radioactive forms. The study of the induced radioactivity gives a sensitive method for the detection of many elements, at levels well beyond the reach of conventional methods.

An element particularly suitable for study in this way is arsenic, once an important instrument of homicide and still an industrial poison in many places. The amount of arsenic in a single hair—about 10^{-10} gm—can be measured by the activation method, and many interesting experiments have already been made to take advantage of this extraordinary sensitivity.

* Shortly to be demolished, according to a recent edict.

Not long ago it was discovered, for example, that two famous detergents contained, along with the fearless cleanser, dazzling whitener and other agreeable constituents, substantial quantities of arsenic. The impurity was found to have entered the manufacturing process in sulphuric acid, made from pyrites by the lead chamber process. The acid contained 0·1 per cent of arsenic and the concentration in the detergent was as high as 70 parts per million—enough to bring the material within the scope of the poisons legislation. The makers, at some financial inconvenience, arranged a fresh source of supply, uncontaminated by arsenic. Samples of the detergents, bought unobtrusively during the succeeding year, showed a steady fall in arsenic concentration, down to insignificant levels.

Arsenic is the only constituent of tobacco smoke known with certainty to be carcinogenic in man. A heavy or moderate smoker increases his daily arsenic intake appreciably. In another experiment done by activation analysis it was thought that the rate of excretion of arsenic (in the hair, skin and nails) might be greater in smokers than in non-smokers. To test this hyopthesis, 1000 samples of hair were obtained from subjects whose smoking habits were recorded, but concealed from the analysts. Using the Harwell reactor, one chemist and a laboratory assistant were able to complete the assay of a batch of 100 samples in two days—a rate that could not have been approached by conventional methods. When all of the samples had been analysed, a correlation was sought between arsenic content and smoking habits. None was found but a strange result emerged from further examination of the figures. It appeared that men's hair contains about 60 per cent more arsenic than women's. This finding still awaits explanation.

The techniques of activation analysis were recently used in solving an ancient historical problem. Erik XIV, King of Sweden from 1560 to 1568, is generally believed to have suffered a disreputable death (in keeping with the rest of his life) but the precise mechanism of his despatch has been a matter of controversy. Activation analysis of a fragment of the royal remains established the presence of abnormal levels of mercury, confirming the legend that he was poisoned, at the instigation of his successor, by mercury concealed in a dish of pea soup.

12. Physics in the hospital service

Many hospital physics departments, in England and other countries, are busily engaged in the measurement and control of radiation, whether as a therapeutic agency, a diagnostic aid or an occupational hazard. A broader concept, now established in Scotland, is that these important tasks form only a part of the physicist's duties in the hospital service. This view has received concrete expression in Glasgow, where the Regional Hospital Board, responsible for some 200 hospitals serving a population of 3 000 000, has contrived a physics department to carry out a wide range of activities. As well as the familiar duties in radiation control, the department undertakes several tasks not usually practised on a large scale in the health service. One group is concerned with the design and construction of scientific instruments used in medical research. Another team is devoted to acoustics, maintaining standards for audiometric testing and carrying out research on the assessment of deafness, particularly in children.

The extent and nature of a deaf person's disability are studied by measuring his auditory threshold at different frequencies, using an audiometer; this instrument comprises an oscillator, an attenuator and an earphone (fig. 4). Though simple in conception it is difficult to maintain in good order; acoustical standards, being largely subjective, are more elusive than standards of length, current, light intensity or any of the other quantities regulated by physicists. Fifty audiometers used in Scottish hospitals were examined after the inauguration of the calibration laboratory in 1957; not one was within acceptable distance of the current British standard. With the co-operation of the manufacturers a greater accuracy has now been achieved, but frequent inspection and overhaul are necessary because no one in the world makes headphones of good enough quality. Every type tested so far shows random variations in sensitivity and frequency response, even over a period of months. The Scottish audiometer calibration service is the only one of its kind in the world. Such a facility would be quite beyond the power of a single hospital to provide but it can be done quite economically on a national basis.

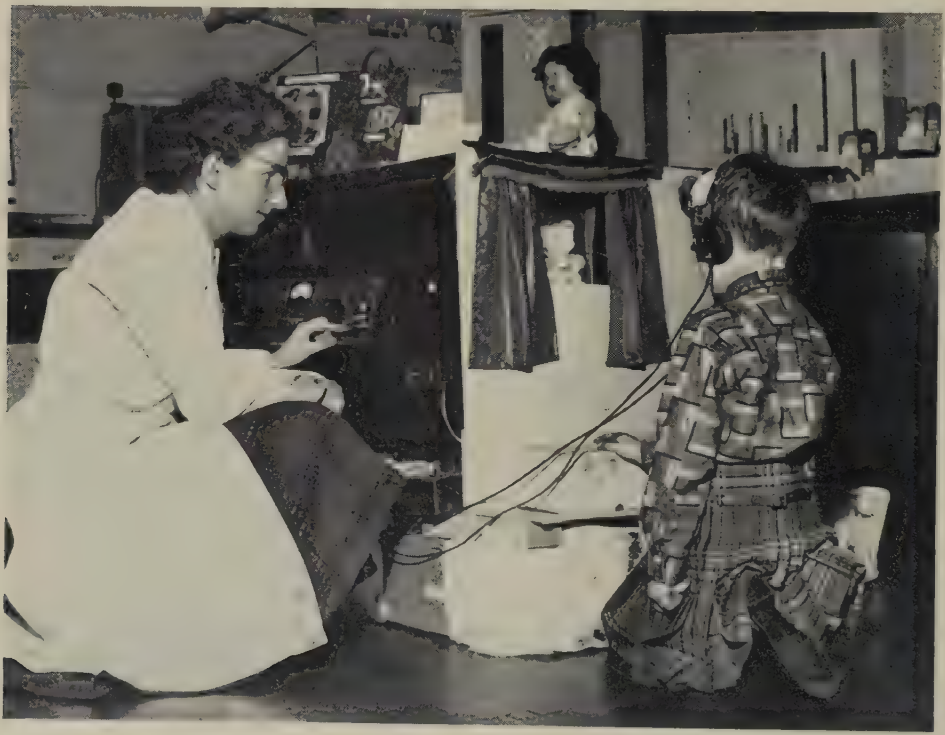


Fig. 4. Measuring auditory thresholds in children. The patient is first conditioned, by the application of loud signals, to press a button when he hears a sound in the headphones; a toy then pops up and gives him a sweet or other reward. The signal is then reduced until the child's hearing threshold is reached. This device is useful down to the age of 3 years and is often the only accurate method of audiometric testing in spastic or mentally retarded children. (*Courtesy of The Glasgow Herald.*)

Another need that the Glasgow department tries to fulfil is the provision of research facilities for physicians or surgeons engaged in problems requiring a lot of scientific help. It is seldom economical for a doctor to engage a physicist for his own unit, particularly if the research in hand is to last for only a few years. A large physics department can, however, supply scientific and technical help for such projects, offering workshops, equipment and other background facilities as needed. The work on respiratory physiology, already mentioned, is organized in this way. The physics department provides a senior physicist and a technician along with all of the necessary apparatus and laboratory facilities. The director of the project—in this case the professor of Midwifery—provides patients, hospital resources and clinical research workers. In this way he secures the requisite scientific and technical aid, at a level that can easily be adjusted to meet special needs, without the added responsibility of conducting his own physics department, a task that is hard enough for a physicist to do and well-nigh impossible for a doctor.

The physics department is busy also with the measurement of radioactive strontium (from fall-out) in children's bones. About a third of all the United Kingdom measurements in 1959 were done by a small team in Glasgow, using the resources of the children's hospital and the Regional Physics Department. This work requires frequent (though not full-time) help and advice from physicists skilled in the more refined forms of radioactive assay as well as from technicians adept at the building and maintenance of electronic devices. It would be difficult (and extravagant) to assemble all of these services, along with the essential chemical skills, in a special unit, but they may readily be obtained as marginal benefits from a substantial physics department.

13. The future of hospital physics

The department devotes a substantial effort to formulating and implementing technical policy on matters of concern to the administrative bodies responsible for conducting the hospital service. A consumers' advisory service for radiologists and hospital boards has established standards for relating the needs of an individual department to the capabilities of the various types of equipment offered by manufacturers. Advice generated in this way has produced striking financial savings without impairing the efficiency of the radiological service—indeed it is often possible, by careful choice of equipment, to provide useful additional facilities without any extra expense. Other problems tackled by the x-ray technology group include the determination of optimum working life for radiological apparatus and the reduction of radiation doses incurred by patients.

These and related activities occupy a team of about 50 people in Glasgow. Schemes of a similar kind have been considered in other places and there is little doubt that many of them will come to fruition, for the physicist has an exciting role in the future development of medicine. The traditional instruments of the physician's craft are as valuable as ever but they have been so successful that many diseases which were once common have now been eliminated or brought under control. There remain some formidable clinical problems in which scientific help on a considerable scale is necessary for diagnosis or treatment. So far this help has come largely from chemists, biochemists and bacteriologists. Remembering that the importance of measurement has been a guiding principle in medicine for three centuries, it would be strange if the huge advances in physics during recent years had not produced ideas and instruments serviceable to the physician. To mention one example, the design of small, simple and rugged scientific instruments for missiles or satellites has much in common with the problems met in trying to measure various bodily parameters on a patient.

14. THE NEW HUMANITY

The foregoing paragraphs illustrate two of the three considerations mentioned at the beginning of this paper as determining the relationship of physics to medicine. The first is the idea that science can be useful.

The second factor is the importance of measurement, which was recognized in medical practice before it became commonplace in the scientific laboratory. The current upsurge of activity in the applications of physics to medicine springs from the availability of more and more refined techniques for measuring quantities of interest to the physician.

The third consideration is more important and illustrates the main point of the present discourse which, the reader will be relieved to know, is about to emerge. Those who read the Scottish newspapers will occasionally notice an advertisement seeking the services of a professor or lecturer in humanity. It may be wondered why a civilized and cultured race like the Scots should need instruction of this sort. To find the explanation we must look back over five centuries to the time when the universities of Scotland were being established. They were all foundations of a medieval kind, in which knowledge was divided into two parts—divinity, the study of God, and humanity, the study of man. Humanity included all the branches of learning that a man might need in his worldly life. Foremost among them was Latin. Though now rightly esteemed as a subject of great cultural value, it was in those days an intensely vocational study for many people, since it opened the door to the professions and the upper reaches of public employment. Furthermore the problems facing the individual (if he was a man of affairs) or the community were often resolved by reference to the literature or philosophy of the ancient world.

In our own time, many of the problems demanding serious attention are scientific in nature. Physics in particular has become a subject of great political, economic and industrial importance: it has indeed ceased to be a speciality and has become a generality. The man who is ignorant of physics cannot fully understand the world in which he lives: still less can he take any informed part in its affairs. Physics, in short, is the new humanity.

It need not surprise us that many of the scholars who have groped their way towards this consummation in days gone by were doctors. Physics and medicine, enjoying a common origin reinforced by centuries of useful co-operation, must today be counted among the most substantial achievements of the human intellect. The bridge between them has been built by men of deep compassion, guided by profound learning—men such as Sanctorius, Krüger, Black, Young and Helmholtz—men, indeed, of great humanity. Against this inspiring background the continued efforts of contemporary doctors and physicists can, though humble in themselves, bring rich rewards in the advancement of learning and the relief of suffering. These are the fruits of the incongruous yet fertile union of the scalpel and the spectroscope.

The Author :

Dr. J. M. A. Lenihan, a graduate of Durham and Glasgow, has been physicist to the Western Regional Hospital Board, Glasgow, since 1953. He is Chairman of the Editorial Board of the British Journal of Applied Physics.

Physics for Grammar Schools

This brief note on Part I of the Science and Education Report issued by the Science Masters' Association and the Association of Women Science Teachers will deal with two points only. Sister St. Joan of Arc's article on page 403 has shown the kind of progress that is being made towards the introduction of new material. We might look at that which is being displaced, and also at the prospect of equipping teachers for presenting modern physics when the time comes.

At Ordinary level, the chief omissions are the parallelogram and triangle of forces, equilibrium of non-parallel forces, kinematics, *sums on* Newton's laws of motion, quantitative magnetism and electrostatics, magnetometry, dip-circle drills, the potentiometer, Wheatstone's bridge, coefficient of expansion of liquids, vapour pressure, photometry, and calculations on curved mirrors and spectacles. Most of these can better be done later by the future scientist, and can well be spared by other pupils.

At Advanced level, suggested omissions are: method of dimensions (for derivation), angular momentum, moment of inertia, n and k, viscosity, diffusion, photometry, refraction at a single spherical surface, Newton's law of cooling, most of sound except the characteristics of wave motion, Gauss's theorem, something called Coulomb's theorem, electrometers, Kirchhoff's laws, and details of thermoelectricity. Here, angular momentum and moment of inertia are fundamental, and needed for the first steps from the mathematician's 'particle' into either the macroscopic world of real moving bodies or the microscopic world of elementary atomic physics; I can see only one reason for rejecting them, and that is that pupils find them very difficult to grasp, though I have never found out why—perhaps it is because they associate them with advanced mathematics. They are equally dim, incidentally, about couples; there might be a case for trying to implant the simple concepts about rotation earlier, since moments are among the few topics remaining in the 'O' level mechanics.

The Report stressed the importance of courses to bring teachers up to date, and an experimental schoolmasters' course held from 10–14 April at the Reactor School, Harwell, through the enterprise of the Director, Dr. F. A. Vick, illustrated what can be done and how it is welcomed. This was not directly related to the proposed new syllabuses, and indeed the organizers made it clear that it was not intended to brief science masters on the kind of thing they ought to teach, but rather to give them the background knowledge that would enable them to approach new tasks with confidence. Telescoped in time, and with the best instructors at the Reactor School working at full stretch throughout it, this was the kind of course normally given to nuclear engineering students. Besides the full lecture programme, four afternoons were spent seeing what goes on at Harwell, and two evening sessions of practical work showed the type of experiment that can be performed by students—given appropriate equipment. Perhaps in due course one of the participants may be able to contribute a short account, and it would be interesting also to hear what the Reactor School concluded from this experiment. This course was heavily oversubscribed; and that so many teachers were keen to attend, and that Harwell was able to provide it, are both promising signs for the future. G. R. N.

On the Record

1. Organization of the Reactor Group of the United Kingdom Atomic Energy Authority

When the Authority's new Reactor Group came into being at the beginning of April, it assumed central control of the largest single organization in the world for the design and development of nuclear reactors. Its facilities at Dounreay, Windscale and Winfrith include full-scale prototype power reactors, zero-energy reactors, critical assemblies, exponential devices and experimental reactors, as well as the ancillary services. The reactors and sub-critical assemblies are :—

NESTOR = NEutron Source Thermal Reactor.
SCORPIO = Sub-critical Carbon mOderated Reactor assembly for Plutonium Investigations.
DIMPLE = DeuterIum Moderated Pile, Low Energy.
SGHW = Steam Generating Heavy Water reactor.
NERO = Sodium (Na) Experimental Reactor of O power.
HELEN = Hydrogenous Exponential Liquid-moderated ExperimeNt.
HECTOR = Heated Experimental Carbon Thermal Oscillator Reactor.
ZENITH = Zero Energy NItrogen-heated THermal reactor.
ZEBRA = Zero Energy Breeder Reactor Assembly.
PLUTO = PLUTOnium loop-testing reactor.
HERO = Hot Experimental Reactor of O power.
DMTR = Dounreay Materials Testing Reactor.
APEX = APproach EXponential.
FIFI = FIne Flux Investigation.

There are also TOAD, TESSIE, PHOENIX, PUMA, SIRIUS, and PANTHER, for which the acrosticians have been so ingenious that one hasn't a clue—or (as they say in the profession) a light.

(*U.K.A.E.A.*, *April* 1961.)

2. Plans for a second Scout Patrol

The British National Committee on Space Research, the U.K. Steering Group on Space Research and the U.S. National Aeronautics and Space Administration have agreed on the scientific instrument payload for the second U.K. Scout Satellite to be launched by NASA. The scientific instruments, as in the first U.K. Scout Satellite, will be designed and provided by U.K. scientists.

It is expected that the satellite will be launched in about two years' time. The orbit, which is expected to range between 300 and 2000 km from the earth, will probably pass over the U.K. and the U.S. and it should be possible to receive radio information both in these countries and at stations in the British Commonwealth.

The experiments planned deal with :—

1. *Galactic Noise*
   Dr. F. Graham Smith (Mullard Radio Astronomy Observatory, Cavendish Laboratory, University of Cambridge) will have three objectives: (*a*) to measure the intensity of the galactic background at as low a frequency as possible (in the range $0{\cdot}75-3$ Mc/s) under known conditions of receiver sensitivity and under known conditions of ionospheric effects on propagation and aerial impedance; (*b*) to watch for temporal or spatial variations of galactic radiation; (*c*) to explore the electron density in the upper F region.

2. *Atmospheric Ozone*
   Dr. R. Frith and Dr. K. H. Stewart (Air Ministry, Meteorological Office, London) will measure the vertical distribution of ozone in the atmosphere as often and in as many places as possible. These measurements should add to our knowledge of the processes forming and destroying ozone, of the air motions which distribute it and of the effects of ozone on the thermal equilibrium of the upper atmosphere. Measurements will be made of the intensity of the radiation received from the sun at selected wavelengths in the ozone absorption region in the ultra-violet at times when the satellite is entering or leaving the earth's shadow and the solar rays have to pass through the earth's atmosphere to reach the satellite.

3. *Micrometeorite Flux*
   Dr. R. C. Jennison (Nuffield Radio Astronomy Laboratories, University of Manchester, Jodrell Bank, Cheshire) will detect optically the holes formed in a thin metallic film by the impact of micrometeorites. Sunlight passing through the holes will fall upon a photosensitive strip of solar cells and give a signal from each hole in turn from which the number and size of the holes, and hence of the micrometeorites, can be deduced. The sensitivity is such that particles 1 micron in diameter or greater may be detected.

(*Joint Press Release from the Office of the Minister for Science and the Royal Society, April* 1961.)

3. NEW DUCTILE SUPERCONDUCTING ALLOYS

New alloys made of molybdenum and technetium were announced at the Spring Meeting of the American Physical Society (April 27) by Dr. B. T. Matthias of Bell Telephone Laboratories, who reported cryogenic experiments showing that these new alloys, which are very ductile, become superconducting at temperatures higher than for any other alloy. (Nb_3Sn, the superconductor with the highest known transition temperature, is a brittle compound.) An alloy of molybdenum-technetium is superconducting at temperatures near 16°K. Technetium is a man-made element, the most stable isotope of which Tc^{99} has a half-life of 2×10^5 years.

A Bell Laboratories team, J. E. Kunzler, E. Buehler, F. S. L. Hsu, and J. H. Wernick, reported in February 1961 that a compound of niobium and tin (Nb_3Sn), at 4·2°K, remained superconducting in fields as large as 88 000 gauss while carrying a current of 150 000 amps/cm^2. Special metallurgical techniques are required to form this brittle alloy into magnet coils that can withstand the mechanical forces produced by very high magnetic fields. Ductile superconducting alloys promise to simplify the problem and will complement Nb_3Sn. Experiments indicate that an electromagnet with about a three-inch layer of windings of the niobium-zirconium alloy will produce a magnetic field of 80 000 gauss at 1·5°K. Results with Nb_3Sn at temperatures between 14°K and 18°K, by extrapolation to lower temperatures, suggest that Nb_3Sn will remain superconducting at temperatures around 4°K, in fields of 200 000 gauss and possibly higher. Therefore Nb_3Sn is still the most promising material for magnets having field strengths well above 100 000 gauss.

A successful superconducting electromagnet will have many important implications. Large magnetic fields will extend the operation of many electronic communication devices to higher frequencies, providing increased bandwidth for use in radio-relay communication systems. Another attractive application is in thermonuclear fusion for the production of electric power. High magnetic fields are needed to provide 'magnetic bottles' to contain high-temperature gas plasmas. A third important use is in the research laboratory where high magnetic fields, at low cost, will facilitate fundamental investigations.

(*Bell Telephone Laboratories, New York, April* 1961.)

4. CARBON-16

Workers at A.W.R.E. Aldermaston, with a team at the Clarendon Laboratory, Oxford, report the production of a fourth carbon isotope by bombarding carbon-14 with 6 MEV tritons, $C^{14}+H^3 \rightarrow C^{16}+p$. The mass of C^{16} is 16·014702 a.m.u., the half-life is 0·74 sec, and it undergoes β-decay with the subsequent emission of a neutron to give N^{15}.

(*U.K.A.E.A., May* 1961.)

5. POLYMERS OF CARBON DIOXIDE

At A.E.R.E. Harwell, commercial carbon dioxide at room temperature and pressures 1–5 atmospheres was being examined by mass spectrometer. The mass spectrum showed ions corresponding to polymers $(CO_2)_n{}^+$. The possibilities that such ions are formed within the mass spectrometer itself or during expansion of the gas from the reservoir into the molecular-beam system have been shown to be slight. It is thus most probable that the ions were formed from polymeric molecules already existing in the carbon dioxide, of the form $(CO_2)_2$, $(CO_2)_3$ etc., continuing in decreasing abundance as far as $(CO_2)_{23}$. This discovery may be important in work on carbon dioxide at high pressures, on carbon dioxide as a coolant, and on the graphite–carbon dioxide reaction in AGR systems.

(*U.K.A.E.A., May* 1961.)

6. THE DENSITY OF MERCURY

In a paper presented on June 8, Dr. A. H. Cook of the National Physical Laboratory described how the density of mercury at 20°C has been measured by weighing the mercury contained in a cube formed of slabs of fused silica, the internal dimensions of which have been measured by optical interference. The results agree among themselves to better than one part in a million and are within one part in a million of the results of complementary measurements using the mass of mercury displaced by a dense cube.

(*The Royal Society, June* 1961.)

A Survey of Research on the Transuranic Elements

by M. B. WALDRON
Metallurgy Division, A.E.R.E., Harwell

1. Introduction

The history of the transuranic elements may perhaps be regarded as commencing with the discovery of nuclear fission on the eve of the last world war. Out of this epoch-making scientific development has come not only the exploitation of a new source of energy for both constructive and destructive purposes, but also the means by which a completely new field has been opened up for the speculation and experimentation of physicists, chemists, mathematicians and applied scientists.

At the same time, the formulation of the periodic table which preceded this development by half a century had provoked questions in laboratory and classroom as to whether elements beyond uranium would be discovered. In point of fact, some elements, neptunium and plutonium, are now known to exist in nature, for example pitchblendes contain about 1 part in 10^{13} parts of plutonium and 1 part in 10^{14} neptunium. But this discovery was not made until more than a year after their production and identification by Seaborg and his co-workers in 1940-41 by neutron bombardment of uranium.

A curious turn of events led to much scepticism when the actual production of the transuranic elements was achieved. The anticipation of such an expansion of the periodic table has led to an erroneous interpretation of the early work of people such as Fermi on the irradiation by slow neutrons of uranium. The radioactivities so produced were later shown to be due to the creation, not of new elements, but of fission products of uranium which were isotopes of existing elements in the middle of the periodic table. Consequently it needed careful studies of the decay pattern and later the oxidation statesof $_{93}\mathrm{Np}^{239}$, for example, to convince the scientific world that the prematurely announced transuranic elements could in fact be produced. After the legality of neptunium was established, plutonium (94), curium (96), americium (95), berkelium (97), californium (98), followed in the space of the next ten years in the order given. There was by this time no doubt that a new series of elements with distinctive properties had been produced. A detailed review of the actinide elements is given by Katz and Seaborg (1957).

To those interested in the fundamental nature of these elements, their behaviour is made the more intriguing by the variety of properties, physical and chemical, that they exhibit. This is most readily seen from the number of chemical valency states that exist or from the high allotropy of elements between thorium (90) and americium. For example, plutonium, as we shall see, has no less that six allotropic forms, ranging from complex crystal structures (e.g. monoclinic) to simple face- and body-centred cubic structures, with widely differing physical properties. This variety of behaviour has made difficult and at the same time fascinating the interpretation of these elements in terms of the periodic classification of the elements.

It is, however, now clear that elements 95 (americium) to element 103—as yet unmade—generally behave as members of a series called actinides since the correspondence of americium to europium and successive elements to the analogous rare earths suggest a series in which actinium must be regarded as the initial member, just as lanthanum is for the lanthanides. Whereas in the lanthanide series, the progressive filling of the 5*d* electron levels or shell is interrupted by the filling of the 4*f* shell, with the actinide elements the filling of the 6d shell is interrupted by the filling of the 5*f* shell. In general, the *f* shells do not extend out as far from the nucleus as the *d* shells and hence do not in themselves react strongly with neighbouring atoms to give strong bonds. The strength of the bond is connected with physical properties such as density and melting point. However, in some circumstances, *d* and *f* levels may become ' hybridised ' to give strong directional bonds which will be referred to later. The correspondence with the lanthanides for those heavy elements between actinium and plutonium is however obscured by complications that will be discussed later. Because of this actinide behaviour, it will be necessary to consider the transuranic elements together with elements 89–92 in this review, if their relationship to the other elements is to be appreciated.

There are several ways in which studies of the transuranic elements differ from all that has gone before. In the first place, the means by which these elements were produced involved that the initial work of all of them had to be on a very small scale. Much of the work on neptunium was done with microgram quantities of the metal or its compounds while in the case of nobelium (102) the identification has been made by elution on a sample containing only a few atoms with a half-life of ten minutes. In latter years it has proved possible to isolate from tons of waste solutions gram quantities of elements such as neptunium, while americium has been produced in similar quantities from highly irradiated plutonium solutions. Because of its importance as a fissile material, plutonium has been produced in kilogram amounts and the size of experimental samples in Britain, Canada, America, France and Russia is limited less by supply than by questions of experimental technique.

Another distinctive feature of work with these elements concerns the radioactivities associated with them. These lead to hazards of various kinds. Where there is mainly α–emission, as with Np^{237}, Pu^{239}, Am^{241} etc., the work has to be done in conditions which prevent the emitters obtaining entrance into the bodies of those handling it. The radiation itself is stopped by the slightest barrier, e.g. a few microns of skin tissue or perspex, a few centimetres of air. If, however, the material is absorbed into the body by inhalation or ingestion, it does intense damage to the body tissue in the immediate vicinity of the emitting particles. In the case of Pu^{239}, which is very slowly excreted from the body and settles in the bone structure destroying blood-making tissue, the limit of plutonium which it is considered safe to retain in the body is set by international agreement at $0{\cdot}3\ \mu g$, equivalent to one moderate-sized dust particle, assuming this was completely retained in the body. For very small quantities of material, say milligram quantities of plutonium, a fumehood may provide sufficient protection. In all other cases it is necessary to work in hermetically sealed enclosures with glass or perspex fronts to provide visibility and with long rubber gloves sealed into special gloveports to enable operations to be carried out by hand within the enclosure (Walton, 1958). Such a device is called a drybox or more correctly a glovebox. In its simplest form it may be constructed of wood and perspex and have an air atmosphere slightly below atmospheric pressure, but in many cases an elaborate construction in mild or stainless steel with a purified inert gas atmosphere and various additional safeguards may be necessary. A typical glovebox used at Harwell is shown in fig. 1. In some cases a number of gloveboxes may be linked together in a chain with a carefully designed layout so that in a relatively small compass one carries out functions normally occupying a whole suite of laboratories. By careful planning it is often possible to offset to some extent the time-consuming complications of glovebox working, once the facilities have been installed.

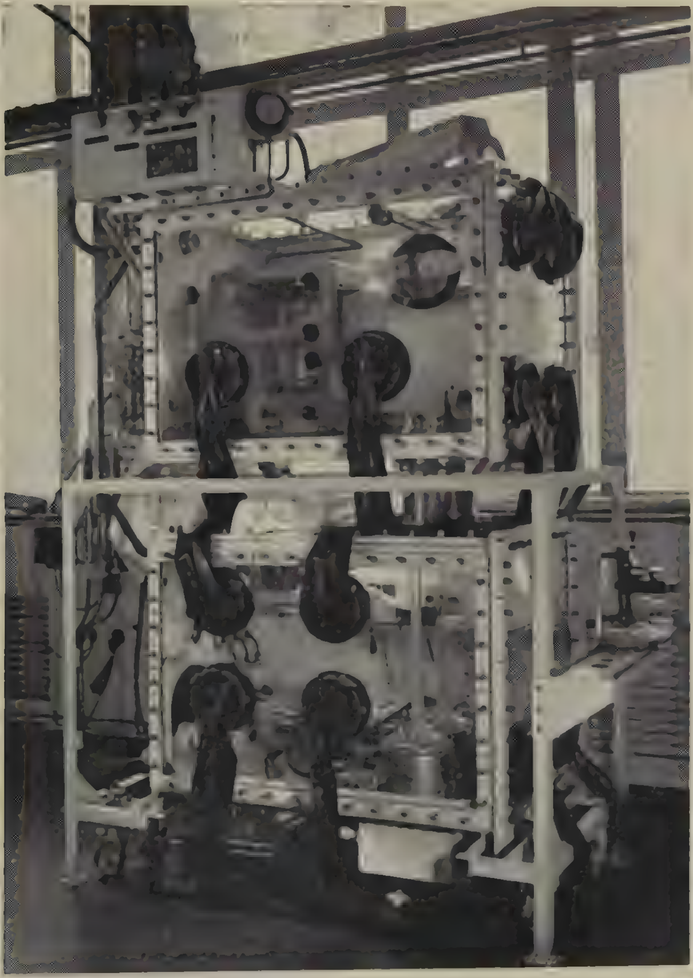


Fig. 1. Typical glovebox used at A.E.R.E.

In many cases there is the added complication of β and γ emission, either from decay of the isotope under examination or because of difficulty in separation from associated elements. In these cases some shielding is necessary, the thickness of which depends on the intensity of the radiation. Again, if the intensity is low, local shielding within the box and the use of specially designed tongs, making use of the inverse square law, may be sufficient. On the other hand, elaborate cells made of special concrete or lead walls up to 1 ft. thick, in which operations are carried out by remote control, may be necessary (Walton 1958). With such isotopes much is to be gained by working with the smallest amounts of material that are practicable.

There is one further effect of importance in studies of these radioactive materials and that is the energy liberated as heat in the disintegration process. Even where the half-life is relatively long, as with Pu^{239}(24 360 y), the heat generation may be sufficient to raise the centre temperature of a specimen a few mm thick by several degrees above ambient and a kg mass by as much as 60°c. In the case of short lived isotopes such as Cm^{242} (162·5d), for which it has been stated that a massive piece of metal would reach incandescence, the problems are very great. Clearly the precise determination of properties of such materials is made difficult or impossible both by temperature effects and by damage to the crystal lattice by the radiation produced. For this reason, the isotope first available in quantity may only be used as a temporary expedient until a more stable isotope can be obtained to take its place. For example, macroquantities of Pu^{242} ($3{\cdot}8 \times 10^5$y) are being impatiently awaited, despite the relative abundance of Pu^{239}.

2. Methods of production

It is not necessary here to go into details of the production and separation of all the transuranic isotopes but it is helpful to mention the principal methods to afford some understanding of the work involved before a study of these elements can begin.

Table 1 lists some of the more interesting isotopes and indicates their modes of formation and disintegration as well as their half-lives. It will be seen that by far the largest number of these isotopes are obtained by long irradiations of Pu^{239} (which itself is formed by neutron irradiation of U^{238}, the isotope which forms 99·3% of naturally occurring uranium). Obtaining these isotopes in sufficient bulk is therefore dependent on the availability of reactor space for such long irradiation experiments together with the careful separation processes involved under conditions of relatively high radioactivity. A second source of some isotopes which depend on decay processes is to extract small amounts from larger quantities of radioactive wastes as for Np^{237}, or to store a source material such as Pu^{241} (or long irradiated plutonium) while the desired isotope, in this case Am^{241}, develops. In such cases the source material concerned may be itself desired, in the meantime, for other experiments which may involve laborious reconversions to other chemical forms at intervals between ‘ milking ’. Alternatively, there is a large capital investment and monopoly of radioactive laboratory space while the ‘ americium-cow ’ or its equivalent, is kept on the shelf. Other methods involve the use of particle accelerators for ion bombardment in which case the production of substantial amounts is not very probable. One further point that is often overlooked, especially in metallurgical studies, is material that is initially ‘ pure ’ to a certain standard, may immediately begin to deteriorate by the production of decay products. For example, the plutonium normally used for experimental work has been highly refined from elements such as iron, aluminium, silicon etc., but since it is a mixture of isotopes including some Pu^{241}, it immediately starts producing month by month tens of

Table I

Isotope	Half life	Mode of disintegration	Mode of formation	Remarks
Np^{237}	$2{\cdot}2 \times 10^6$y	α	$U^{237}\beta^-$ decay	Useful for macro work, 100g quantities separated.
Np^{239}	2·33 d	β^-	$U^{239}\beta^-$ decay	Tracer work, 1st element identified.
Pu^{239}	24 360 y	α	$Np^{239}\beta^-$ decay	Fissile element available in tens of kg.
Pu^{242}	$3{\cdot}8 \times 10^5$y	α	succ. n γ from Pu^{239}	Available in μg quantities.
Pu^{244}	$7{\cdot}6 \times 10^7$y		succ. n γ from Pu^{239}	Not yet available.
Am^{241}	458y	α, β^- stable	$Pu^{241}\beta^-$ decay	Forms from long irradiated Pu.
Am^{243}	7600y		succ. n γ from Am^{241}	Available in mg quantities.
Cm^{242}	162·5 d	α	Pu^{239} (α, n)	High self heat.
Cm^{244}	19y	α β stable	succ. n γ Pu^{242}	g quantities available U.S.
Cm^{245}	$1{\cdot}4 \times 10^4$y	α β stable	succ. n γ Pu^{242}	μg quantities available U.S.
Cm^{247}	4×10^7y	α β stable	succ. n γ Pu^{242}	} Not yet available.
Cm^{248}	$4{\cdot}7 \times 10^5$y	α β stable	succ. n γ Pu^{242}	
Bk^{247}	7×10^3	α	Cf^{247} E.C. decay	Awaits advanced deuteron and helium ion accelerators, very desirable.
Bk^{249}	290d	β^-, α	succ. n γ from Pu^{239} with β^- decays	Expected to be available in μg quantities eventually.
Cf^{249}	470d	α β stable	Bk^{249} β^- decay	} Small amounts made by intense neutron irradiation of Pu^{239}
Cf^{251}	660y	α	Pu^{239} succ. n γ with β^- decay	
E^{254}	280d	α	succ. n γ from Pu^{239} with β^- decays	Desirable.

parts per million (a significant level for many impurities) of americium and the effect of this is not really known although in most respects it is probably not very important.

The chemical separation of the various isotopes is a very extensive field and depends greatly on the circumstances and the elements concerned. In general, two stages of separation are involved where a fissionable isotope has preceded the end product in the formation route, firstly removal of fission products which range around the middle periods of the Periodic Table and in general have a distinctive chemistry, and secondly, separation from the other actinide elements. Oxidation and reduction using oxalates, peroxides, phosphates and halides are commonly used for precipitation routes, while solvent extraction methods using hexone, tributyl phosphate etc. are also very powerful, both groups depending on the marked differences of behaviour of the elements in their various oxidation states, i.e. the valencies exhibited by the metals in their stable compounds. A given oxidising condition may lead to different oxidation states for the different elements, with more marked variations in properties than would exist between compounds of the same type and valency. Ion exchange resins are very useful for separation from other actinide and rare earth elements.

For physical and metallurgical studies on metallic behaviour, including alloy studies, the further step of reduction to metal is necessary. Since these elements are highly reactive and readily form stable oxides, nitrides and carbides this step is a critical one especially where small quantities of metal are involved. In general, most methods depend on reduction of an oxide, chloride, or fluoride of the metal concerned by reactive metals which are sufficiently different in their atomic structure to resist alloying with the metal concerned. Barium, calcium, magnesium and beryllium are among the reductants commonly used. The process involves heating the finely divided compound and reductant, intimately mixed, to a temperature at which the exothermic reaction produces molten globules of metal clearly separated from the other materials. If the heat of reaction is insufficient to raise the metal to its melting point, the metallic powder formed may have to be leached out chemically, the large surface area of the powder product making it very difficult to avoid oxidation. It is sometimes necessary to add a booster such as iodine to provide additional heat of reaction where the quantities are small. The reactivity of the metals poses further problems in the choice of container material and crucibles, since reaction with these at the reduction temperature can seriously contaminate the metal produced. Owing to the high atomic weights of the actinide elements, a few parts per million of the higher impurities corresponds to a much higher proportion by atoms of the impurities, the latter being the more significant figure when the effects on properties are under consideration.

Having briefly outlined the historical and experimental background of the study of the transuranic elements, we will consider what is known of them individually from a physical point of view. Since this information is very incomplete and complex, some attempt will be made subsequently to review the overall behaviour of the elements as a series and to discuss the areas in which progress can be currently expected.

3. PROPERTIES

3.1. *Thorium and Uranium*

It is necessary to review the properties of these metals because they form the introduction to the pattern of behaviour that develops through the transuranic series (Wilhelm and Rodgers 1955, Bell 1957, Wilhelm 1958, Grainger 1958, Holden 1958). The principal properties for all the actinide elements are given in table 2 from which this trend is apparent. For example, the melting point of thorium is 1750°c and is well above that of actinium (1050 ± 50°c) which precedes it, fig. 2. This would seem to be a reflection on the existence of two electrons in the 6*d* shell with an associated high cohesive energy as is seen in the transitional elements generally. As we pass through protoactinium about

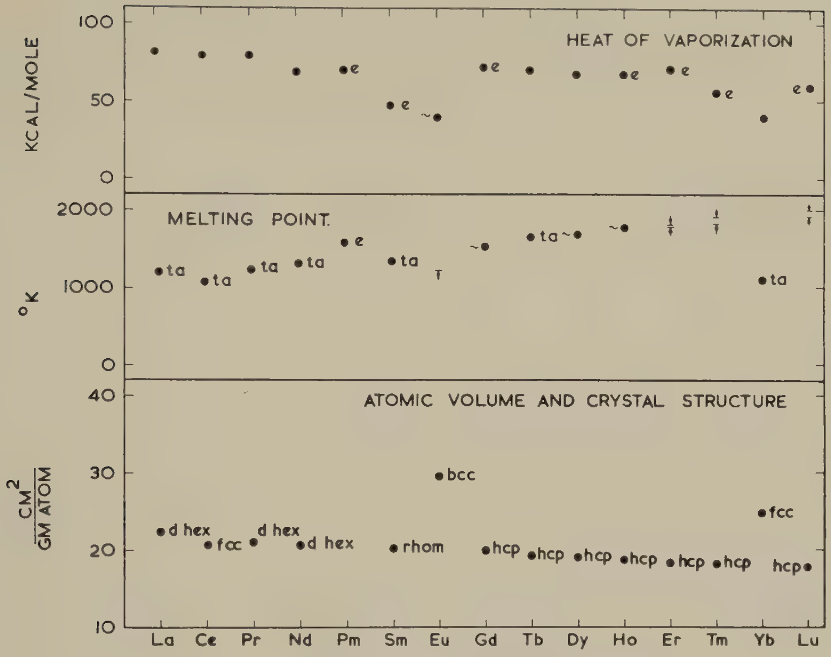


FIG. 2.(a) LANTHANIDE METALS.

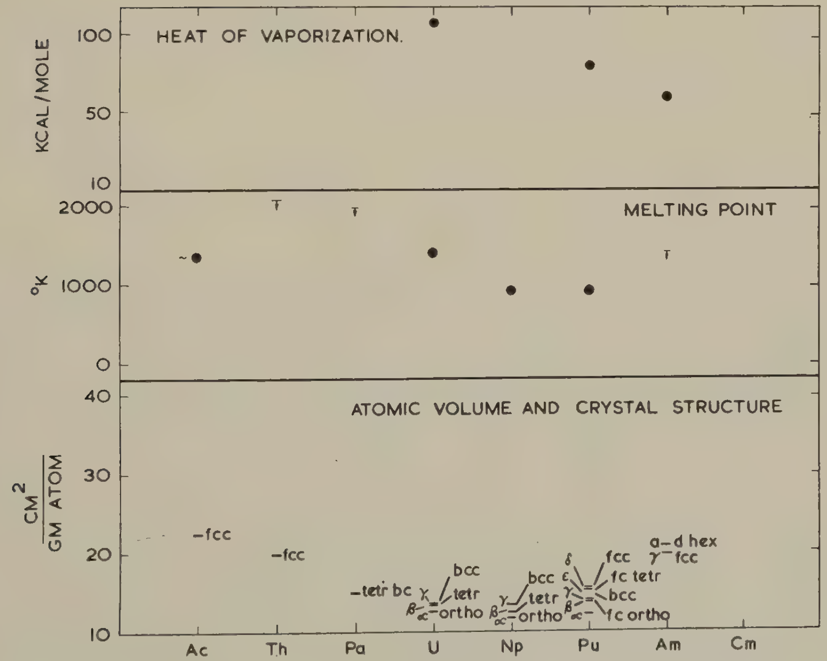


FIG. 2.(b) ACTINIDE METALS.

Fig. 2. Comparison of heats of vaporization, melting points, atomic volumes and crystal structures of lanthanide and actinide elements (Seaborg and Katz, 1957).

which little is known, to uranium the melting point drops to 1132°c with one 6*d* electron and the filling of 5*f* levels which have sufficient spatial extension to participate in bonding although in a more lanthanide-like manner. This melting point is intermediate to the low values (around 600°c) found in the transuranic elements.

Table 2

	Melting Point °c	Density g/cc	Expansion Coefficient m/m°c.10^6	Resistivity (ρ) Ω.cm.10^6	$\frac{1}{\rho}\frac{d\rho}{dt}\cdot 10^5$	Crystal Structure
Actinium	~1050	~10	—	—	—	f.c. cubic
Thorium	~1690					
α		11·72	+12	13	+396	f.c. cubic
β		11·00	—	—	—	b.c. cubic
Protoactinium	<1873	15·37	—	—	—	tetragonal
Uranium	1130					
α		19	+19	~30	+50–500	orthorhombic
β		18·1	+22	54	+18	tetragonal
γ		17·9	+18	50	+73	b.c. cubic
Neptunium	639					
α		20·45	+27·5	114	+0–46	orthorhombic
β		19·36	+41	107	+0–31	tetragonal
γ		17·90	+33	109	−6	b.c. cubic
Plutonium	639·5					
α		19·82	+55	145	−21	monoclinic
β		17·80	+40	110·5	−6	b.c. monoclinic
γ		17·14	+35	110	−5	orthorhombic
δ		15·92	−9	103	+7	f.c. cubic
δ'		16·00	−16	105	+45	f.c. tetragonal
ε		16·51	+36	114	+7	b.c. cubic
Americium	<1100 (830?)					
α		11·87	—	—		double hexagonal c.p.

The unusual prevalence of allotropy becomes evident with thorium, which has both a body centred cubic structure at temperatures above 1400°c and a face centred cubic structure at lower temperatures. Uranium likewise has a body centred cubic structure above 773°c and two complex crystal structures at lower temperatures. It will be seen that similar low temperature complex structures occur with neptunium and plutonium. The densities of the various phases of thorium and uranium are interesting since the increase in density between the respective elements from about 11 to about 19 precedes high values for neptunium and the low temperature forms of plutonium, values around 20 being obtained which are only exceeded by elements in the platinum group.

The increasing density through the early members of the series reflects a progressive reduction in the atomic radius of the metals concerned. This variation has a considerable influence on many metallurgical properties, particularly the alloying behaviour since only atoms which are close to each other in size will replace each other in metallic solid solutions and compounds. It is perhaps easier to visualise the connection between atomic size and the spatial distribution of the electrons or the cohesive energy of the atom.

The complex crystal structures such as those of α-and β-uranium are associated with anisotropy of properties measured along the principal crystallographic axes of single crystals or strongly textured polycrystalline samples in which a predominance of specific crystallographic orientations in particular dimensions can be detected and estimated. This effect is particularly notable in the case of the expansion coefficients of α–uranium in which there are large positive coefficients for the *a*– and *c*– directions but a negative coefficient for the *b*–direction. This behaviour appears to be related to the curious crystal structure of α–uranium in which there are corrugated sheets of atoms separated by exceptionally long bonds in the *b*–direction. As a result of this anisotropy, an individual grain, on heating, will endeavour to expand in two directions and contract in the third. When surrounded by grains of differing orientation internal stresses will develop which will cause deformations between grains until the strains are accommodated. Unless the uranium is finegrained and randomly oriented, there will be overall distortions (growth) and surface irregularities (wrinkling) whenever the uranium is heated and cooled. These effects are of considerable importance in nuclear reactor technology.

The specific physical properties of thorium, protoactinium and uranium are not in themselves very remarkable and will be referred to when general trends through the series are discussed. In the cubic phases all their properties are typically metallic, having shiny surfaces when clean, being low but not abnormal conductors of electricity, having hardness, tensile strength and ductility comparable with other metals, particularly those of the first long period. They are strongly paramagnetic and form compounds with ferromagnetic, as well as paramagnetic characteristics. The directional nature of the α– and β–uranium structures introduce anisotropy as has already been indicated but even so the mean values are not particularly unusual. The specific heats of thorium and uranium are much higher than the classical value attributed to lattice vibrations and suggests that there is a high electronic contribution in all these elements, as will be discussed later.

3.2. *Plutonium*

It is convenient to discuss the behaviour of plutonium next, partly because it stands out most clearly from all elements that both precede and succeed it in the actinide series and also because, by an accident of birth, the technological importance of plutonium has led to a more complete survey of its occurrence and properties, (Coffinberry and Waldron 1955). We can then take a backward look to neptunium before considering the later members of the series about which little is known.

Plutonium has at the same time a very low melting point and the largest number of allotropic modifications among all the known elements. There are six stable crystallographic forms in the range 0—640°c. The three forms which exist at higher temperatures are again typically metallic phases but the lower three have complex structures with directional properties and some resemblance to the lower temperature phases of uranium. In addition to the exceptionally large number of phases, a glance at the density values in table 2 will show that there is a remarkable variation in the density of the metal from phase to phase. First of all the overall variation from 19·82 at room temperature, a very high value suggestive of transitional metal behaviour once more, to 15·92 in the δ phase, which is a very wide range. Secondly the density decreases as far as the δ phase and then increases in the higher temperature δ′ and ε phases. These features are seen in fig. 3 which shows the expansion behaviour of plutonium up to its melting point. Both the high degree of allotropy and the density variation can be attributed to the special state of the electron energies for the 6*d* and 5*f* levels at this point in the series as we will see later. Depending on circumstances, by which we mean temperature, atomic environment, etc., the plutonium atom can choose a ‘ *d* ’ type (or transitional type) electronic state or an ‘ *f* ’ type (or lanthanide rare earth) state. This choice affects the atomic volume, that is to say the density, on the one hand and the type of bonding, or lattice geometry, on the other. In the case of plutonium this choice can evidently be exercised in several different ways without making much difference to the energy or stability of the structure. It is a corollary of this that the relative stability of the phases will be readily affected by other variables, such as the introduction of foreign atoms, impurities or alloy additions, which will introduce electronic or other factors of a different kind. It is in fact the case that some of these phases are quickly eliminated in favour of others by very small amounts of such additions ; for example, the δ′ phase is eliminated by only a few hundred p.p.m. of aluminium or silicon.

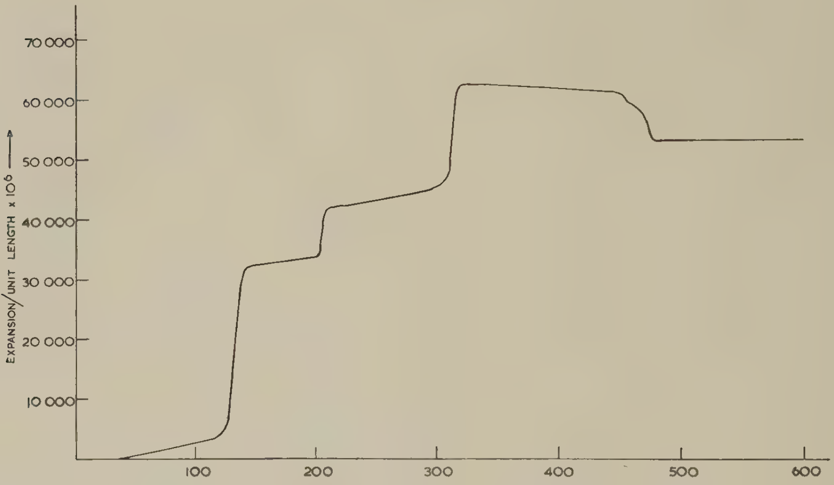


Fig. 3. Expansion curve of pure plutonium (Lee).

There is one further item of special interest concerning the dilatometric behaviour of plutonium on heating, namely the negative coefficient, or contraction of the δ and δ′ phases. The δ phase is face-centred-cubic, an isotropic structure, and the δ′ is similar except for a very slight tetragonal distortion. A contraction in some directions of an anisotropic substance is not uncommon (e.g. α–uranium, already cited) but this behaviour is believed to be unique among simple isotropic structures. Several reasons have been advanced for this, none of which are entirely satisfactory, but they generally suppose that there is a progressive duality of electronic behaviour with increase of temperature resulting in an efficient contraction through transitions from 5*f* to 6*d* states. One possible line of explanation for example is that as the temperature is increased within the δ–phase, there is a change from 5*f* to 6*d* character of more and more electrons giving a negative coefficient of expansion until the energy relations are such that the structure prefers to change to the slightly anisotropic tetragonal structure of δ′. The reason for the change to delta–prime could be possibly due to an ordering of atoms with electrons in two different configurations, which presents certain theoretical difficulties, or to general lowering of the energy in this way by an optimising of the interaction between the electrons and the periodic structure of the lattice, a so-called Brillouin Zone effect.

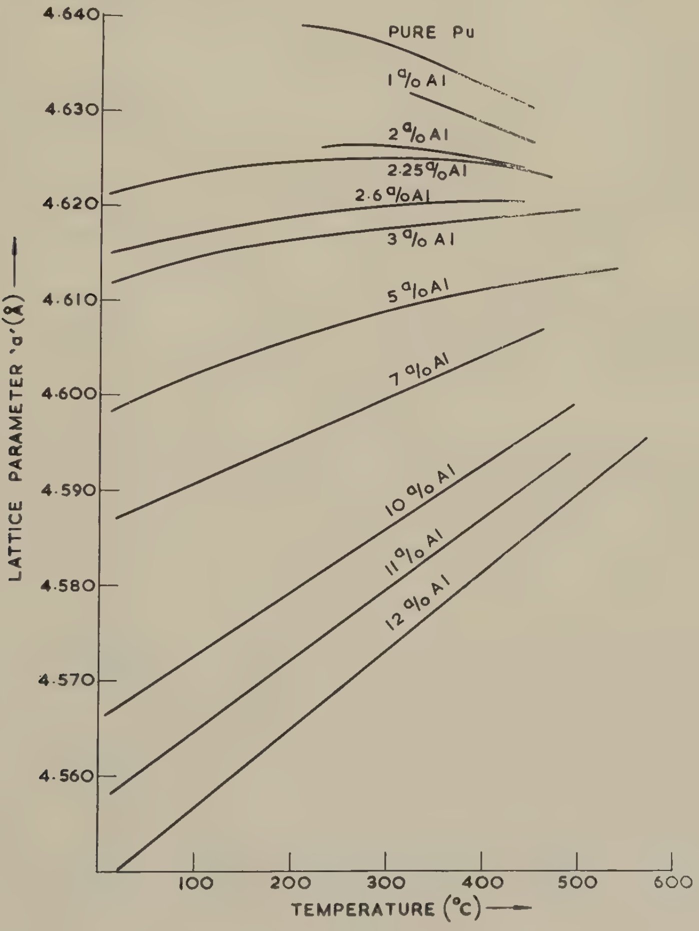


Fig. 4. Expansion curves for plutonium-aluminium alloys (Lee).

Further indications of the special nature of the phases of plutonium come from the electrical properties. The absolute value of the resistivity in the various phases is high but not unduly so in relation to the rest of the periodic table. It is however true that the resistivity coefficients are anomalous in some of the phases such that the resistivity and dilatometric coefficients are not both normal in any one phase. The significance of this is not understood but is the kind of behaviour that might be associated with Brillouin Zone or atomic ordering effects. One explanation for negative coefficients of isotropic materials has been proposed, (Varley 1956), which does not involve any change of electronic configuration. An approximate calculation shows that a suitable combination of electronic energy and Brillouin Zone characteristics could produce such an effect. However there are other physical measurements in the case of plutonium which appear to be inconsistent with this.

It is fortunate that there are a number of elements which dissolve extensively in the δ phase of plutonium without changing its crystal structure. In doing so, most of them extend the range of temperature over which the delta structure is thermodynamically stable while in practice atomic mobility often becomes so sluggish that the phase can be cooled down to low temperatures without any change in structure taking place. Such alloys are called ' delta-stabilised ', and they enable the effects of changes in electron concentration, the average number of electrons per atom, to be studied, since by adding atoms of varying ' valency ' in varying proportions to plutonium the ' concentration ' can be systematically varied. Lee and his co-workers at Harwell in association with Meaden and Mendelsohn of the Clarendon Laboratory, Oxford, are systematically studying the properties of these alloys with aluminium, cerium and zirconium, for example, both at elevated temperatures where these anomalous effects are observed, and at temperatures down to less than 1°K where electronic effects can more readily be distinguished from effects due to the thermal vibrational energy of the lattice. Figure 4 shows the high temperature expansion behaviour of a series of plutonium-aluminium alloys in which the effect of aluminium in changing the negative coefficient of expansion of the pure δ phase into positive values is seen. By measuring resistivity, thermo-electric power, magnetic susceptibility and dilatometric changes on such a series of alloys from near absolute zero to the upper temperature limit of the δ phase, various aspects of the electronic energy contribution and bond formation should be deduced.

3.3. *Neptunium*

Until recently the small quantities available of neptunium, which is an α–emitter similar to plutonium, together with the technological priority attaching to the latter element, resulted in little work being done with it. However, enough information has become available to show how neptunium fits into the actinide series pattern of behaviour.

The low melting point of neptunium marks the end of the drop in melting point from the high value of the thorium, with a relative constancy along the remainder of the series whose melting points are known. Its allotropic behaviour is similar to that of uranium, having three crystalline forms including a body centred cubic high temperature γ phase, although no precise correspondence exists between the lower anisotropic structures of uranium and neptunium. The density values are all very high and compare with those of uranium, indicating the continuance of the strong bonding characteristic of d–type transitional metal behaviour. There is nothing notable about the expansion behaviour of neptunium but the resistivity in the γ phase has a temperature dependence which suggests anomalies in the electronic behaviour of neptunium analogous to those already discussed for plutonium. This behaviour is also found in the γ phase in certain uranium alloy systems. The specific heat of neptunium is even higher than those of its near neighbours which are themselves in excess of the classical value attributable to lattice vibrations. This suggests an exceptional electronic contribution consistent with the complex electronic energy levels proposed for these elements.

A limited amount of research into the alloying behaviour of neptunium has been undertaken particularly with its neighbours uranium and plutonium. These are notable in having shown extensive inter-solubility of these elements in the lower temperature phases having complex crystal structures. Normally, such complex structures do not tolerate more than a fraction of one per cent of alloy addition in the lattice, this being attributed to the quasi-chemical nature of the bonds as indicated by bond angle and bond length observations. These bond characteristics are explicable in terms of mixing *s p* and *d* ' orbitals ' in particular ways. It therefore appears that neptunium can enter into these bond formations readily and hence can give rise to the extensive solubilities that are observed. For example, neptunium dissolves up to about 50% in both orthorhombic α-uranium and monoclinic α-plutonium, whereas most other elements are virtually insoluble in these phases. It is also interesting that the bond arrangements around individual atoms are very similar in some of these structures, even though the complete lattices fall into different crystallographic classes.

As with plutonium, measurements of carefully chosen physical properties over the whole temperature range are being undertaken by Lee and co-workers in order to elucidate the electronic and bonding behaviour.

3.4 *Americium*

Knowledge concerning americium metal is very limited indeed, mainly because of its scarcity and because of the short half-life of Am^{241} which is the isotope available in mg or g quantities. It is hoped that Am^{243} will be available for metallurgical studies before too long.

The metal is said to be more silvery in appearance than neptunium or plutonium and to be very ductile in comparison. Some doubt exists concerning the melting point which is certainly below 1100°c and possibly below 800°c. Thus it may be somewhat higher than neptunium or plutonium although this cannot be regarded as proven. The density is low, about 11·8 g/cc at room temperature, in the α phase which has a double hexagonal crystal structure similar to that of europium in the lanthanide series. There are said to be two other phases of americium at higher temperatures but no particulars of these are available.

It will be seen from these observations that americium is the first element in the actinide series to show in its metallic form true rare-earth behaviour similar to the lanthanides. The low density and associated large atomic diameter, together with the unusual hexagonal crystal structure which only occurs in lanthanide elements, all mark it off from the preceding members of the series.

3.5. *Curium and the Trans-curium elements*

Very little is known about the physical properties of the remaining actinide elements. One would expect them to be low melting point metals, moderately ductile and of moderate density and some preliminary observations of beads of curium more or less support this. Further work will certainly await milligram or gram supplies of longer-lixed isotopes.

4. The general picture

The most complete understanding of the behaviour of the transuranic elements as a series comes from the chemical studies. The oxidation states are shown in table 3, in comparison with the data for the lanthanide rare-earths. It will be noted that whereas the lanthanides have normally only 3–valent states, the earlier actinides show a wide variety of oxidation states among which the 3–valent state does not become predominant until americium. The variety of states in the earlier elements indicates the indeterminate energy relations between the 5*f* and 6*d* electron levels in the early members of the series. This can be understood more readily from fig. 5. Associated with the increasing stability of the 3–valent state, increasing difficulty of oxidation as measured by oxidation potentials is noted.

Table 3

Oxidation states of Actinide and Lanthanide elements (Seaborg and Katz, 1957).

Lanthanide Elements:										
Atomic number	57	58	59	60	61	62	63	64	65	66
Element	La	Ce	Pr	Nd	Pm	Sm	Eu	Gd	Tb	Dy
Oxidation states	3	3 4	3 (4)	3	3	2 3	2 3	3	3 (4)	3
Actinide Elements:										
Atomic number	89	90	91	92	93	94	95	96	97	98
Element	Ac	Th	Pa	U	Np	Pu	Am	Cm	Bk	Cf
Oxidation states	$\underline{3}$	(3) $\underline{4}$	(3) 4 $\underline{5}$	3 4 5 $\underline{6}$	3 4 $\underline{5}$ 6	3 $\underline{4}$ 5 6	$\underline{3}$ (4) 5 6	$\underline{3}$ (4)	$\underline{3}$ 4	$\underline{3}$

The correspondence between the later lanthanides and actinides can also be clearly shown by studies of the ion-exchange resin absorption and elution behaviour. Figure 6 shows a typical set of elution results in which the correspondence is very marked. It is also seen from the ionic radii of the actinide elements in their compounds that there is a contraction analogous to the ‘ lanthanide contraction ’ as you pass along the series, figure 7. This is a consequence of the addition of successive electrons to the inner 5*f* shells causing an increased charge on the nucleus of the atom. Magnetic measurements on such ions also show the close similarity between the series but since the interpretation of the data is both involved and controversial, more detailed comment is unjustified here.

When we move from the chemical evidence to the physical properties of the elements themselves which have been discussed in some detail already, the picture, though less clear, remains substantially unchanged. As we pass from thorium to neptunium there are many similarities to transitional elements in which electrons are assigned to the ‘ d ’ orbitals. Thus melting points of the early members are high, densities are high, particularly in the low temperature modifications and the crystal structures of these modifications can be generally understood by considering the atoms as entering into quasi-chemical bonds involving 5*f* and 6*d* orbitals, combined to give specially strong directed bonds with well defined bond angles.

An attempt has been made (Friedel, 1957) to generalize this trend, as illustrated by fig. 8, by suggesting a progressive transition from pure '*d*' orbital behaviour at lower temperatures in the early members, through '*d–f*' hybridized or combined orbital behaviour in the vicinity of plutonium, particularly at higher temperatures, and finally '*f*' orbital or lanthanide behaviour in americium and later elements. In agreement with this idea of a gradual progression, we can note the reduced stability of the body centred cubic temperature phase favoured by transitional metals, and the development of many allotropic modifications in the intermediate zone. We have already seen how much of the detailed physical properties also could be connected with ease of interchange between the different electronic states and we can note the high density of the lower modifications of plutonium giving way to the lower density forms nearer the melting point. In understanding these changes it is important to see that as the '*d*' and '*f*' level energies approach equality, as seen in fig. 5, the difference in energy becomes comparable with the difference in binding energy of different compounds and atomic arrangements. Hence the distribution of the electrons between energy levels may vary for a given element from one allotropic phase to another or from compound to compound.

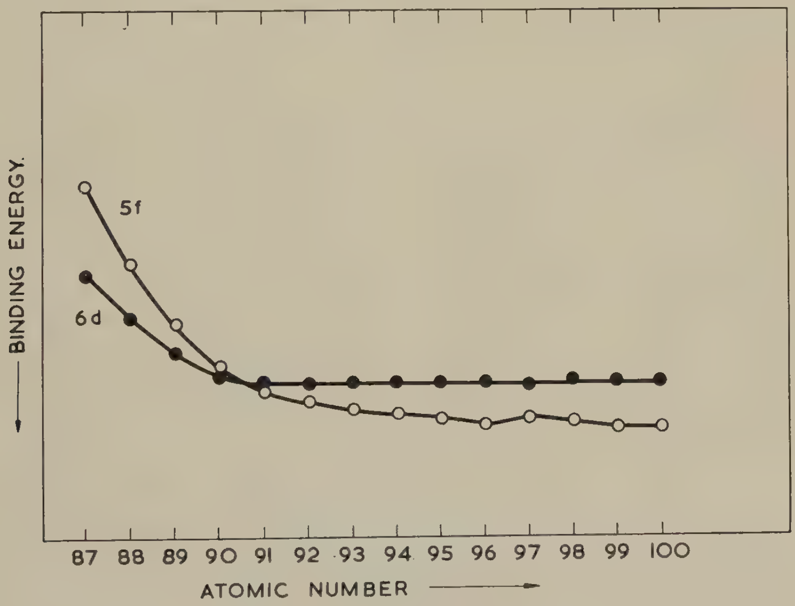


FIG. 5.

Fig. 5. Qualitative representation of electronic binding energies in the heaviest elements (Seaborg and Katz, 1957).

5. The periodic table

The variability of behaviour of the earliest transuranic elements and the increasing tendency of electrons to occupy the 5*f* shell with increasing atomic number make it difficult to arrive at a satisfactory arrangement of these elements in the periodic table. The scheme most commonly adopted emphasises the actinide anology to the lanthanides and shows them as an interpolated series immediately below the lanthanide rare-earths. If this presentation is followed it is desirable to note very plainly that thorium, protoactinium and uranium have a typical behaviour. Various authorities have endeavoured to represent their behaviour by showing either a correspondence of these elements to hafnium, tantalum and tungsten, with neptunium and its successors corresponding to the rare earths from praseodymium onwards, or even by showing a thoride series under hafnium, containing the transthoride elements. Devices of this nature, however, appear to lay too much emphasis on the special behaviour of the earlier members.

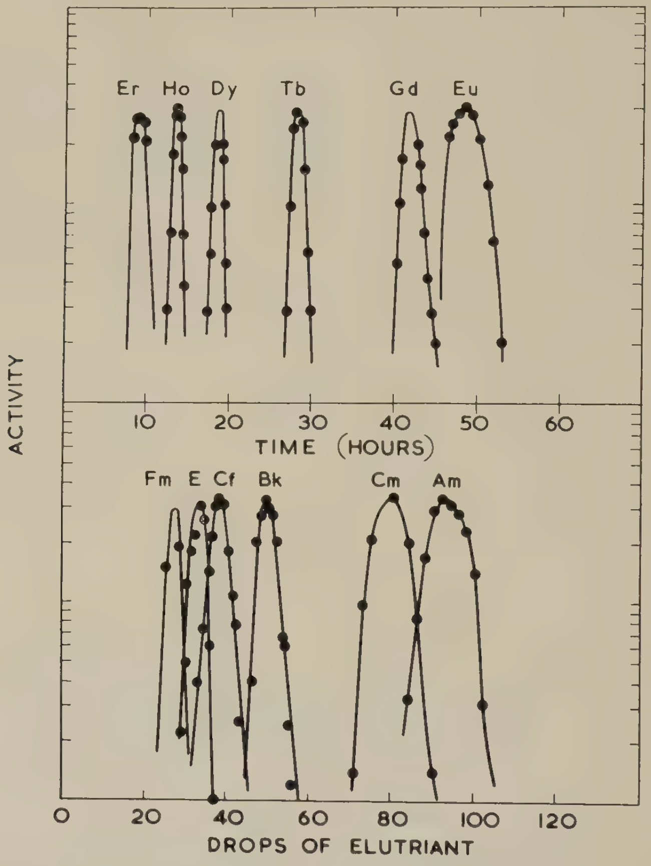


Fig. 6. Elution of tripositive lanthanide ions and actinide ions from Dowex-50 cation exchange resin with ammonium citrate element (Seaborg and Katz, 1957).

6. Future developments

Although the general picture can be stated with some confidence in the way we have done, much remains to be done before the elements themselves can be understood in detail. For this reason, the concerted study of neptunium and plutonium by means of selected physical properties is most interesting. It is encouraging to know that such work is now being revived in American laboratories and being taken up actively by French workers, following up the interesting results obtained by Lee at Harwell. As long as the quantities are limited, results are always liable to be influenced by impurities since it is usually more difficult to eliminate these on small-scale operations. We therefore look forward

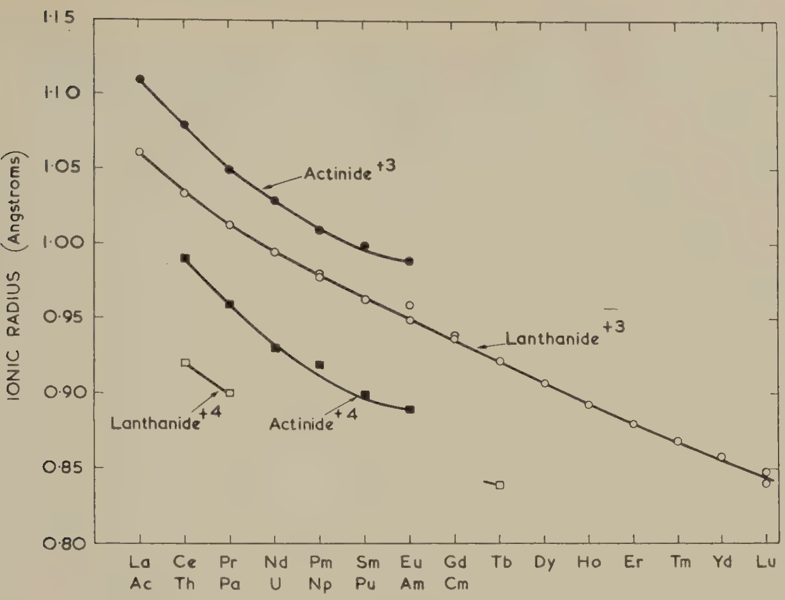


Fig. 7. Ionic radii of actinide and lanthanide elements (Seaborg and Katz, 1957).

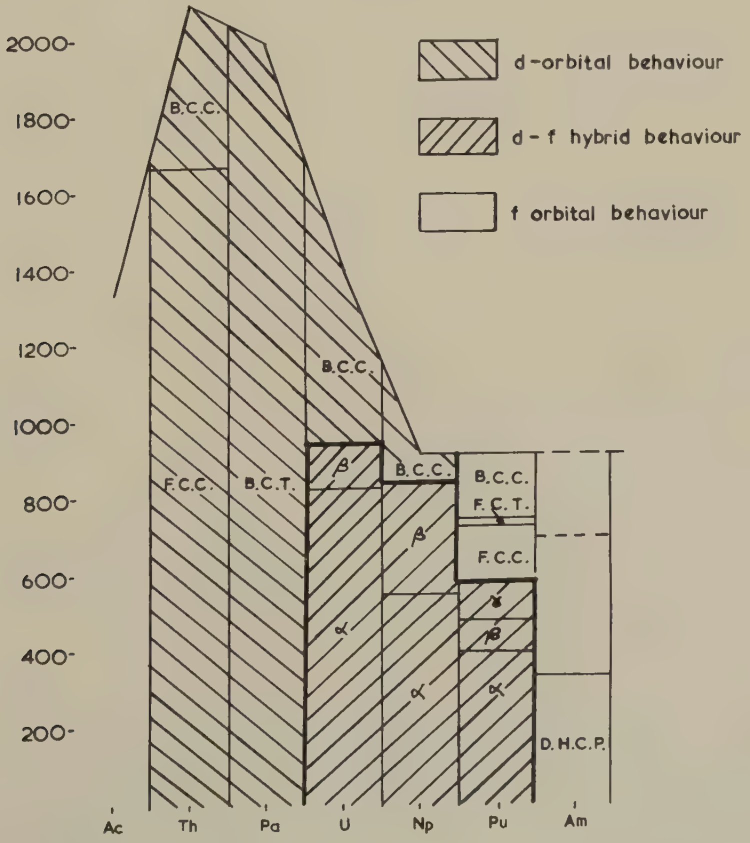


Fig. 8. Melting Points (°K) and crystal structures of the heavy elements (after Friedel, 1957).

to more generous supplies of americium and curium and even of actinium and protoactinium which have been little studied in the past. It has been pointed out that many of the isotopes in use at present are relatively short lived, giving rise to α, β, and γ radiation. Not only is this a hindrance to those who handle these materials but there is often serious doubt on the extent to which the properties themselves are affected by the disintegration, whether from the damage caused by the radiation or from the ‘ foreign ’ atoms produced in the lattice. Consequently we must also hope that the more desirable long lived isotopes like Pu^{240} and Am^{243} may become available for study in these larger amounts. We can safely say that there is a wide scope for future work on every feature of the transuranic elements.

Unfortunately, perhaps, progress in this field is dominated by the practical needs of the nuclear energy industry, since the facilities required are not usually available outside the laboratories set up by the various atomic energy organisations. For this reason the bulk of the work will be confined to plutonium since large quantities of Pu^{239} will be produced each year in the world's reactors, and there is a strong incentive to find the best way of using this byproduct as an alternative fuel to U^{235}. Since Pu^{239} produces a substantial excess of neutrons over those needed to sustain the chain reaction when neutrons are used at or near fission energies, plutonium technology is of particular importance for fast reactors, such as the one now working at Dounreay. The fundamental studies of plutonium lie on a trail whose meandering path coincides sufficiently often with the technological high road to ensure fairly rapid progress in our knowledge of this metal, but for the transuranic elements generally we are dependent on the generosity with which effort is permitted to be diverted in the interests of pure science.

REFERENCES

(1) Bell, I. P., 1957, *Thorium. Nuclear Engineering*, **2**, 19, 418.
(2) Coffinberry, A. and Waldron, M. B., 1956, *Physical Metallurgy of Plutonium. Progress in Nuclear Energy*, **1**, V, 354.
(3) Friedel, J., 1957, *Symposium on Rare Metals*, pp 386–401, Indian Institute of Metals.
(4) Grainger, L., 1958, *Uranium and Thorium*. (Newnes: London)
(5) Holden, A. N., 1958, *Physical Metallurgy of Uranium*. (U.S.A.E.C./Addison Wesley Publishing Co.: Reading, Mass.)
(6) Katz, J. J. and Seaborg, G. T., 1957, *The Chemistry of the Actinide Elements*. (Methuen & Co. Ltd., London).
(7) Lee, J. A., 1961, *Physical Metallurgy of Neptunium. Progress in Nuclear Energy*, **3**. V (to be published).
(8) Varley, J. H. O., 1956, Proceedings of the Royal Society. (**A**), **237**, 413.
(9) Walton, G. N., 1958, *Gloveboxes and Shielded Cells*. (Butterworths.)
(10) Wilhelm, H. A., 1958, *The Metal Thorium*. (A.S.M.)
(11) Wilhelm, H. A. and Rodgers, B. A., 1955, *Physical Metallurgy of Thorium. Nuclear Metallurgy Symposium*. (A.I.M.E.)

The Author:

Dr. M. B. Waldron, B.Sc., Ph.D., was educated at Whitgift School, Croydon, and King's College, London University. He did post-graduate Research at Birmingham University under Professor G. V. Raynor on the alloying behaviour of transitional metals, 1946–1949. He has been associated with metallurgical research on plutonium and neptunium at Harwell since 1953 and is now Leader of the Physical Metallurgy Group in the Metallurgy Division at A.E.R.E. He has fostered collaboration with the other British, American and French laboratories in this field and is particularly interested in securing a proper understanding of the metallic behaviour of the Actinide elements.

The Teaching of Modern Physics in Schools

by

SISTER ST. JOAN OF ARC,
La Retraite High School, Clifton, Bristol.

Summary

A residential study group organised by Dr. L. R. B. Elton at Battersea College of Technology from April 4th–8th considered the practical possibilities of teaching modern physics in the middle forms of grammar schools as proposed in the new syllabuses of the Science Masters' Association and the Association of Women Science Teachers. This article gives an account of the meeting.

1. Background

Everyone concerned with the teaching of science in schools will have noticed in recent years a two-way development of syllabuses. There has been constant pressure for the inclusion of new material and this has been accompanied, though with a noticeable time-lag, by a cry for the excision of old material. The Advisory Council on Scientific Policy went so far as to say (*Annual Report* 1959–60) that 20 to 25 per cent of the science curricula could be removed without harm and indeed with benefit—though those immediately concerned in the matter might have felt happier if the dead wood had been identified before its percentage was quoted. Efforts in both directions had met with no great success. The Science Panel of the Secondary Schools Examinations Council, through whose hands the science syllabuses of the Examining Boards pass for approval, had consistently opposed the inclusion of such topics as ' elementary descriptive treatment of the atom ', ' electrons, protons, neutrons ' on the grounds that it had not yet been shown that such topics could be dealt with scientifically (i.e. in the light of experimental evidence) at the O Level stage, and such examination questions as had been produced on them implied and were likely to encourage a form of teaching which was authoritarian and dogmatic and therefore opposed to the true interests of science teaching.

It became apparent that something more was needed than this piecemeal approach to syllabus reform. The Science Masters' Association and the Association of Women Science Teachers had been working since 1957 on a full Report and in the course of 1960 produced the drafts of a Revised Policy Statement and of syllabuses compiled by their biology, chemistry and physics panels. These were circulated for comment and the final versions of this Part I of the Report prepared. Since both the physics and chemistry syllabuses contained significant amounts of modern physical science and certain Examining Boards were prepared to implement their recommendations as soon as possible, the Science Panel of the S.S.E.C. invited members of the two Associations to a conference at Studley to discuss the new syllabuses. At this conference arose the idea of the Pilot Scheme in O Level physics, notice of which was circulated to schools last February. A number of schools are to be invited to enter pupils for an examination in physics at O Level in summer 1963 or 1964, which will be based on the S.M.A./A.W.S.T. syllabus (including the section on atomic structure and nuclear energy). Not more than two or three of the Examining Boards are likely to be involved; the experiment will be directed by the two Associations and the S.S.E.C. with the assistance ot the Ministry.

2. The Battersea Study Group

Even before the Studley Conference the Radioactivity Committee of the S.M.A. had had its terms of reference enlarged and had grown into the Modern Physical Sciences Committee, and Dr. L. R. B. Elton of Battersea College of Technology had obtained a Nuffield Foundation grant to cover the expenses of a study group to work out a possible presentation of the subject in schools. Some twenty-five teachers from schools, training colleges and institutes of education, together with observers from the Ministry of Education and the London County Council were invited to meet at Battersea in Easter week for a residential study group. The problem facing them was two-fold: a method of approach had to be worked out, and suitable experiments designed both for demonstration and class work.

3. Practical Work

There were three marked influences on the experimental work:

(1) The course for an Arts VIth form worked out by Mr. J. L. Lewis at Malvern College and described to some extent in his ‘ Shell ’ lecture of April, 1960.

(2) The first-class apparatus for the teaching of modern physics developed by Leybold of Cologne.

(3) Apparatus and ideas from the P.S.S.C. course and other U.S.A. sources.

Radioactivity tended to dominate the picture. There were several reasons for this, the most obvious being that this part of the syllabus had interested Mr. Lewis and been worked out by him fairly extensively at Malvern. An initial lecture from Miss B. E. Stern of Battersea College on safety precautions in the handling of radioactive materials gave us the recognised rules for work of this sort, even though it involves nothing so dangerous as crossing the road ! The wise and co-operative attitude of the Ministry in this matter was stressed throughout. Much work in schools can be done with freely available uranium and thorium compounds and when it becomes clear what other sources are necessary in school laboratories, regulations are expected to be amended accordingly. Mr. R. A. Faires of the Isotope School. A.E.R.E., demonstrated experiments with uranium and thorium ‘ cows ’ and other work without the need of even a laboratory coat. It is interesting that this section of the syllabus is at the moment the most developed since at first sight it is the one which immediately raises the questions ‘ Can we? ’ and ‘ May we? ’. The challenge has evidently called for some reaction and it now appears that this is the easiest for class work with simple apparatus. Analogue experiments, with coins or dice to show the statistical nature of the process, or with water flow from tubes to illustrate decay, half-life and radioactive equilibrium, need to be used with circumspection or there is a risk either that the analogy is more complicated than the original process, or that it obscures it. ‘ Sir, this experiment with the radioactive dice . . .’

Although monitors, ratemeters and scalers were used for many experiments, some of the work could be done with the simple gold-leaf electroscope, but two developments of this were found much more useful and practical: the Landsverk electrometer and Leybold's Wulf electrometer. The Landsverk electrometer, used in the Oak Ridge Institute for Nuclear Studies in Tennessee, is basically a quartz-fibre pocket dosemeter similar to our Civil Defence instrument, with an extra long electrode which penetrates to the ionization chamber below; it can be charged by an internal transistor converter device; it costs $125 and with added customs duty is an impossible proposition for schools as it stands. But the principle here seemed sound. We must first of all decide what experiments are necessary and useful and what apparatus is suitable; afterwards it is hoped that it will be possible to design simplified versions and persuade manufacturers in this country to produce instruments in quantity at a reasonable price. This, indeed, is a point at which we might usefully learn from the Russian methods which Mr. Lewis has recently seen. A full-time staff at the Institute of Pedagogy is concerned with the design of suitable apparatus, an order for mass-production is given to industry and the equipment distributed to schools. Perhaps a little less freedom and considerably more help for science teachers in this country would make it possible for the indifferent teacher to give good interesting lessons, and the competent teacher would less often be frustrated in his demonstrations from sheer lack of time and money. It would be an advantage, too, if through this apparatus manufacturers could be persuaded of the general principle that the polished mahogany base is not the essence of good equipment, and if some of the simple effective apparatus designed, for example, by Her Majesty's Inspectors or exhibited by S.M.A. members, could become generally and inexpensively available.

Two methods for the measurement of e/m were investigated; the Leybold fine-beam tube and the English version of the P.S.S.C. experiment which uses an electronic ' eye ', and there was a standard piece of apparatus for Millikan's determination of e. These would be demonstrated at O Level and become class experiments in the VIth form, but it was thought that although Millikan's apparatus should be available it would be too long and difficult an experiment for completion by the VIth form and would have to be supplemented by film or other means. Some doubts arose about the logical validity of the P.S.S.C simplification of the experiment by using plastic drops of standard size, but there is more work to be done here. Mr. T. A. M. Peacocke of St. John's School, Leatherhead, demonstrated (besides some interesting chemistry) a diffusion cloud chamber which had possibilities for schools; in fact, it has been known to work rather better in a litre beaker. Mr. Llowarch of the London Institute of Education has a bicycle pump expansion chamber (not shown at Battersea) which works very well. Mr. M. S. Smith of King's College School, Wimbledon, explained and demonstrated the field-emission microscope. This exciting piece of apparatus, used with a small Van de Graaf generator, shows the pattern of the crystal structure of its tungsten cathode on a fluorescent screen, the magnification being of the order of a million. Temperature movements of foreign atoms on the top are clearly visible. We learnt with some amusement that no company had been willing to insure it against a week's use by ' thirty competent physicists '! There were considerable differences of opinion about its possible use in school.

The weakness in this aspect of the work was the very uneven coverage of the experiments. Certain sections of the syllabus are almost entirely neglected, so there is a great deal more work to be done, but an encouraging start has been made.

4. TEACHING APPROACH

The approach to the subject has still not been adequately thought out. Two lectures from Dr. Elton helped to clarify the problem and suggested a possible line of action, and much discussion, both formal and informal, throughout the week explored the ground. It was generally agreed that there should be no *a priori* assumption that the historical approach was the best, still less the only one; and a sound practical contribution came from Mr. Llowarch on the first part of the syllabus. It was practical not merely in the sense that the apparatus was designed and used, but that a coherent sequence of ideas was put forward. It is possible, he suggested, to pass fairly simply from elementary electrostatics to thermionic emission by eliciting the suggestion that, in view of the kinetic theory, it might be possible by heating to get this separated electric charge from matter into space. Various questions about the current flow in the valve circuit followed by experiment lead to further enquiries which call for the invention of the C.R. tube and a measurement of e/m. Discharge tubes and positive rays would follow. If the rest of the syllabus can be worked out in a way which is equally imaginative and yet down to earth, all will be well.

The biggest immediate difficulty comes with the chemists' evidence for the existence of atoms. When the draft syllabuses were distributed it was strongly urged that co-operation between the physics and chemistry panels was essential. This co-operation has occurred to some extent but has now, it seems to me, reached the stage where chemist and physicist are bowing politely before the same door ' After you!' and neither can progress till a decision is taken. Historically, at least, the basic evidence for the atomic theory lies in 19th century chemistry and it seems a pity that the chemistry panel has rejected its logical development as too lengthy and difficult for the intermediate phase. Work has been done and is being done on this in various quarters and might repay careful study. It is, however, possible that some other scheme which is logically watertight and which does not involve the historical structure of 19th century chemistry could be worked out and prove acceptable to both chemist and physicist. The solution of this problem does seem essential if there is not to be a serious discontinuity, to put it at its mildest, in our science teaching.

The parts of the syllabus dealing with the nucleus and nuclear energy received scant attention for obvious reasons. As someone aptly remarked in discussion " At the mention of the neutron we all rose and went to lunch ". Various incidental problems were considered during discussion periods. What precisely was meant by ' physics for all '? Could the E-stream be expected to enjoy the same dogma-free intellectual exercise as the A-stream? It was felt that the basic structural problem had better be solved first, and the E-stream given special consideration later. Would a different approach be necessary with girls? The idea was unanimously rejected. What balance would be achieved between demonstration and class work? What use could and should be made of film? Some attention was given to films and slides during the meeting; in general it was felt that a number of useful slides were available but there would be a need for newly made films on parts of the syllabus.

5. Why Teach Modern Physics in School?

One's reasons for wishing to include a particular topic in a syllabus are bound to have a weighty effect on one's treatment of that topic. Dr. Elton considered the question early in the week and rapidly dismissed a set of reasons which, however large they loom in many people's minds, can only be called utilitarian. It is true that a certain knowledge of modern science may be counted a necessity for life in the modern world; it may be regarded as self-evident that those who are to direct the policies of governments, whether at the level of the ballot box or of Downing Street, must have at least sufficient understanding to be able to ask the right questions of scientists and evaluate their answers. None the less, if science is to be regarded as a vehicle of education, it must produce stronger reasons than these to justify its position. Science, as a whole, is at no loss to justify its claim to be a major activity of the human mind, an aspect of the great search for truth, and modern physics can be no exception. Why, then, must any further question be asked in its regard? If modern physics is to be included in this general claim of science, then it must be taught in a way that does not make nonsense of it as a creative intellectual activity (as, indeed, must all the rest of the syllabus, though this is not always realised) and the problem then becomes: Is it possible to teach this topic in this way in school, even at the intermediate stage? Many experienced teachers believe that it can be done, and on the experimental side at least, the evidence is growing.

Dr. Elton maintained that modern physics had been responsible for new attitudes and methods in science which could not be understood in the classical context, and this constituted a major reason for its inclusion in the syllabus. This idea seems to be linked with a minor controversy of the moment concerning the use of the plural when one speaks of scientific method. No one, I think, would be prepared to maintain that the method is identical in its application to different fields of enquiry, yet on the other hand, if one is to speak of science as a separate discipline at all, there must be something distinctive in its functioning which justifies the use of the singular term ' method '. Can it then be said that modern physics has brought a change in scientific method, or is it merely that a new field of study has called for different methods of application of what is basically the same tool? Something less drastic than the first suggestion, but deeper than the second seems to be involved: a change in the metaphysical pre-suppositions of science since the last century. The recognition that such pre-suppositions exist is in itself a change that has come with the new physics; the mechanical view of the universe and its accompanying determinism had little room for philosophy, perhaps did not even admit the initial act of faith (it can be no more) that the scientist makes in the uniformity of nature. The scientist of our day may be a wiser and a humbler man than his predecessor, he is less sure of himself, theories succeed each other more rapidly nowadays, and he is less open to the risk of regarding any as ultimate truth. He is perhaps more disposed to open the door to the philosopher when his tools need sharpening, and to recognize both the limits of his own terrain and the limitations of his own method. If the study of modern physics in schools can help our pupils to an understanding of the nature of science as its study has done for those responsible for its growth, then let us do all we can to bring it about. But let us make sure that we do so in a way which will achieve our aims, and not frustrate them by allowing dogmatic statement to replace experimental evidence.

Some sound useful work was done at Battersea, a few extravagances such as a discussion on suitable shielding for our neutron sources were quickly rejected, and plans for continuing the work in small local groups are already being implemented.

REFERENCES

The Annual Report of the Advisory Council on Scientific Policy, 1959–60. (Cmnd. 1167, H.M.S.O., 1960.) 1*s*. 9*d*.

Science and Education (Policy Statement and Syllabuses). Report issued by the Science Masters' Association and the Association of Women Science Teachers. (John Murray, 1961.) Available from the S.M.A. Librarian, 52 Bateman Street, Cambridge. 1*s*. (statement), 2*s*. (syllabus).

Mr. J. L. LEWIS. Lecture on 'The place of Atomic and Nuclear Physics in Modern Teaching' delivered 7 April, 1960, at the 'Shell' course for science teachers. Copies available on request to the S.M.A. Librarian.

Physical Science Study Committee—Physics text. Published in England by Harrap and Co. Ltd., 1961. 4 parts. 40*s*. Laboratory Manual 10*s*.

Articles on the Physical Science Study Committee in *Contemporary Physics*:
  Mr. J. L. LEWIS. (1) A Survey of its Publications, Vol. 1, No. 3, February 1960.
  Dr. H. F. BOULIND. (2) The Summer Institute Courses. Vol. 1, No. 4, April 1960.
  Mr. S. WHITE. (3) The Planning and Structure of the Course, Vol. 2, No. 1, October 1960.

The Author:

Sister St. Joan of Arc was educated at La Retraite and at the University of Bristol. She was elected an Associate of the Institute of Physics in 1958 and is Senior Science Mistress at La Retraite High School, Bristol.

The Applications of Single Crystals

by A. E. DE BARR

Machine Tool Industry Research Association

INTRODUCTION

The word ‘ crystal ’ comes from a Greek word meaning clear ice which was also applied to rock crystal (clear, transparent quartz). It was not until about the 17th century that the word was extended to other naturally occurring, transparent minerals which showed a regularity of outward form. It was appreciated at this time that this regularity resulted from an inner symmetry of atomic arrangement and it has since been realized that almost all solid materials are, to some extent at least, crystalline although this is often not apparent from their outward appearance. Indeed most materials of practical importance are polycrystalline—they consist of an aggregate of randomly-arranged small crystals—although single crystals have been used for special purposes for some considerable time, e.g. the crystals of calcite that are used to make Nicol prisms, and the crystals of galena (lead sulphide) that were used in early ‘ crystal ’ wireless receivers. But in recent years increasing use has been made of single crystals for a wide variety of purposes and the demand for single crystals of particular materials is such that large quantities of single crystals of materials such as ethylene diamine tartrate or terphenyl are now grown artificially by one of several techniques now available for the growing of single crystals.

This article sets out to explain why the properties of single crystals differ from those of the more usual polycrystalline materials and why particular single crystals are required for the various purposes for which they are now used.

1. ANISOTROPY

Crystal symmetry

The atoms of a crystal are arranged in a regular manner around the points of a three-dimensional lattice which has the property that the surroundings of any one point are exactly the same as those of any other point. In the interior of a perfect crystal of sodium chloride, for example, every sodium atom has exactly the same arrangement of sodium and chlorine atoms around it. But it is only when a crystal has been able to grow in a completely undisturbed way that its outward form is able to reproduce its internal symmetry. The crystals of materials of industrial importance usually grow under conditions such that their growth is impeded by that of neighbouring crystals or by the walls of a container, or are subjected after crystallization to processes such as drawing or rolling. Although the ordered atomic arrangement characteristic of the material still remains within each individual crystallite, there is usually little or no long-range order in the polycrystalline material as a whole. Chemically, the polycrystalline material and the single crystal may be identical but their physical properties differ in many important respects primarily because the lattice extends undisturbed throughout the whole volume of a single crystal.

In amorphous materials all physical properties are independent of the direction in the material along which they are measured ; such materials are said to be isotropic. In crystals, however, because the atomic arrangement varies with direction in the crystal, so also, in general, do the physical properties of the crystals, and crystals are said to be anisotropic. The physical properties of the crystal must show the symmetry of its atomic arrangement which, in turn, depends upon the symmetry of the lattice.

There are fourteen different kinds of three-dimensional lattice, i.e. there are fourteen ways in which points can be arranged in space so that each point has identical surroundings. Based on these fourteen space lattices there are seven different crystal systems (fig. 1) and although there is a very much larger number of distinct crystal structures, all crystals belong to one of these seven systems. To each crystal system there corresponds a certain minimum of symmetry, i.e. a certain minimum number of symmetry elements—centre of symmetry, axes of symmetry and planes of symmetry. Individual crystals belonging to a system may have more symmetry than this but they cannot have less. Thus, for example, all orthorhombic crystals must have two two-fold axes of symmetry at right angles to each other, but orthorhombic crystals can also have a centre of symmetry or planes of symmetry. (Rotation by $360°/n$ about an n-fold axis of symmetry brings the crystal into a position exactly equivalent to its original position.) Any given physical property—electrical conductivity, elasticity, etc.—may be of higher symmetry (i.e. more nearly isotropic) than that of the crystal system but cannot be of lower symmetry. In cubic crystals, for example, electrical conductivity or thermal expansion are isotropic—i.e. independent of the direction along which they are measured. Young's modulus of elasticity, however, varies with direction in the manner shown in fig. 2, the distance, in any direction, from the centre to the surface of the solid figure being proportional to the modulus in that direction. The ratio of the greatest and least values of Young's modulus is 1·0 for tungsten but 2·85 for copper, for example, although both metals form cubic crystals. The less symmetrical the crystal system, the less regular is the figure representing the variation of any physical property.

In polycrystalline aggregates the anisotropy of the individual crystallites is usually masked by their more or less random arrangement within the aggregate. Many polycrystalline materials therefore appear to be isotropic, the properties of the material in any direction being a kind of average over all directions of the properties of a single crystal of the material. Thus the usually quoted value of Young's modulus for iron is about $2{\cdot}10^{12}$ dyn. cm^{-2} : in a single crystal of iron the value varies from $2{\cdot}84$ to $1{\cdot}32\times10^{12}$ dyn. cm^{-2} according to the direction of measurement.

A rather special kind of anisotropy is found in crystals that do not have a centre of symmetry, i.e. which are not symmetrical with respect to a point. In such crystals a suitable external force—mechanical, electrical or magnetic—can polarize the crystal and cause opposite ends of one piece of crystal to develop different properties. It will be clear that in a randomly arranged polycrystalline aggregate the average value of the polarization in any direction will be zero and, in general, these single crystal properties cannot be reproduced in polycrystalline materials.

Amongst crystals that do not have a centre of symmetry there is a still smaller group of polar crystals or crystals in which one particular direction—the polar axis—has unique properties. Whereas all crystals that do not have a centre of symmetry can be polarized in the direction of an applied stress, polar crystals can be spontaneously polarized along their polar axis or can be polarized along this direction by some isotropic stimulus such as heat.

Bearing jewels

Bearing jewels in watches and instruments—particularly kilowatt-hour meters—were originally made from natural agate or sapphire, the hardness and smooth finish of which provide a long-lasting, low-friction surface. Most bearing jewels are now cut from single-crystal ' boules ' of artificially-grown sapphire (a form of alumina). The effective life of meter jewels has been found

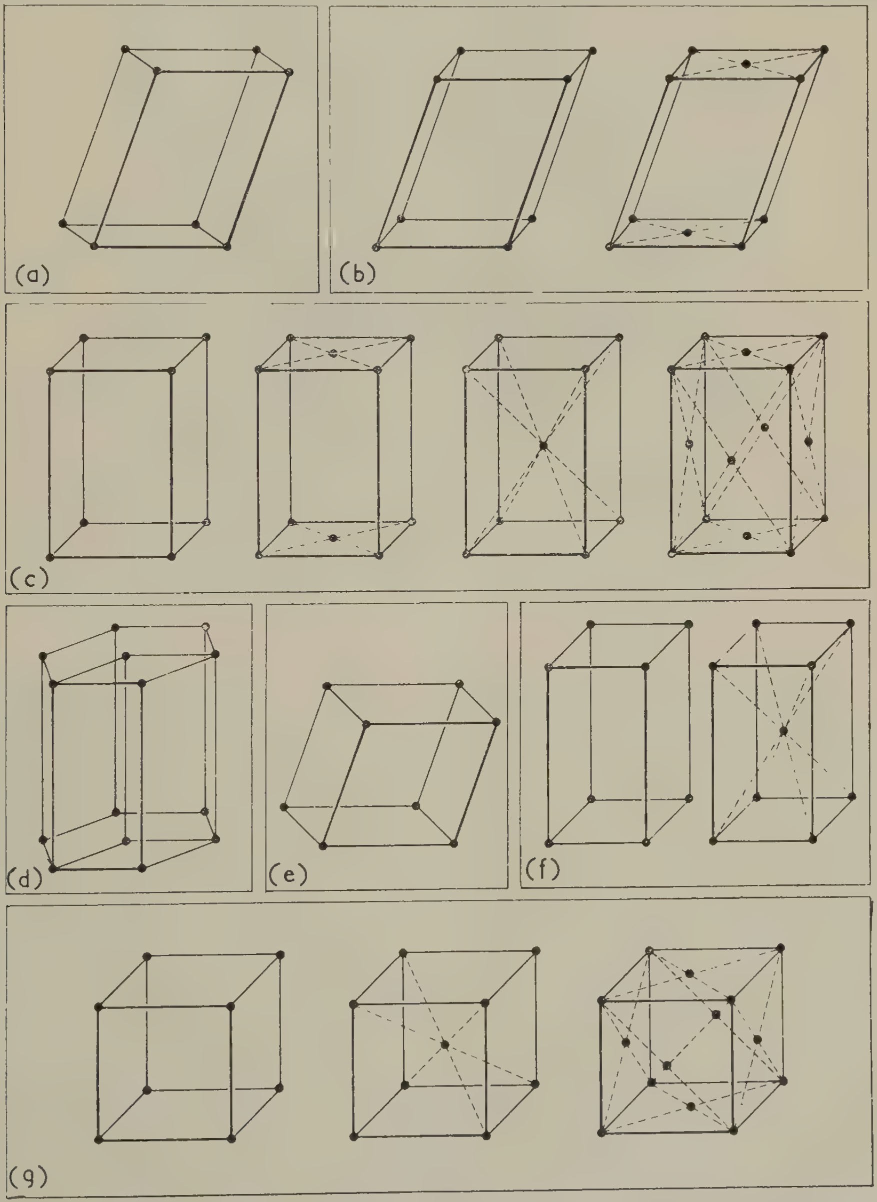


Fig. 1. The seven crystal systems and the fourteen space lattices. (*a*) Triclinic, (*b*) Monoclinic, (*c*) Orthorhombic, (*d*) Hexagonal, (*e*) Trigonal (rhombohedral) (*f*) Tetragonal, (*g*) Cubic.

to depend markedly upon the direction in the crystal of the axis of the jewel. Sapphire has a hexagonal axis (see fig. 1) and the best jewels are obtained when the axis of the jewel is perpendicular to the hexagonal axis. Although there is no complete theoretical explanation of this result, it is associated with the tendency of sapphire to split in particular directions and with the anisotropy of mechanical properties to be expected in a crystalline material.

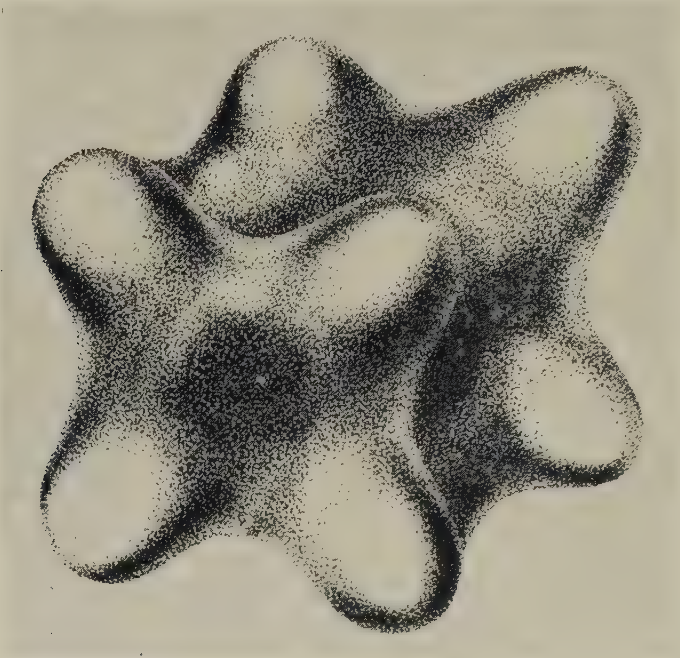


Fig. 2. Anisotropy of Young's modulus of elasticity in a cubic crystal.

Piezoelectricity

In 1880 Pierre and Jacques Curie discovered that when a crystal of quartz or tourmaline was loaded an electric charge proportional to the applied force developed on the surface of the crystal. This is the piezoelectric effect and it is shown by many crystalline substances. The converse piezoelectric effect—the change in the dimensions of a crystal when a voltage is applied across it—was discovered in 1881, and it is convenient to discuss piezoelectricity in terms of the converse effect.

When a voltage is applied across a crystal there may be some movement of conduction electrons producing a current through the crystal but, if this is not sufficiently great to prevent any appreciable voltage remaining across the crystal, there will also be an additional effect—a polarization of the ions making up the crystal. In order to polarize an ion the effective centres of the positive and negative parts of the ion must be separated and, since work is required to effect this separation, energy is stored in the material in the form of electrical polarization. The increase in the energy that can be stored in a condenser by putting some substance between the plates is a measure of the dielectric constant of the substance.

With piezoelectric materials an applied electric field can produce not only polarization of the individual ions but also a change in the separation of the positively and negatively charged ions making up the crystal, an effect which also allows energy to be stored in the material and contributes to its dielectric constant. The change in the separation of the ions represents a change in the dimensions of the crystal : the order of magnitude of the effect is about 10^{-5} %/v per cm.

With very few exceptions all crystals without a centre of symmetry are piezoelectric. Piezoelectricity is essentially a single crystal property since, in a randomly arranged polycrystalline aggregate, the charges produced (or, in the case of the converse effect, the displacements) would cancel out. For a substance to be useful as a piezoelectric material its crystals must also have a high electrical resistivity so that large potential differences can be maintained. The crystals must be reasonably stable and the piezoelectric effect large. For some purposes

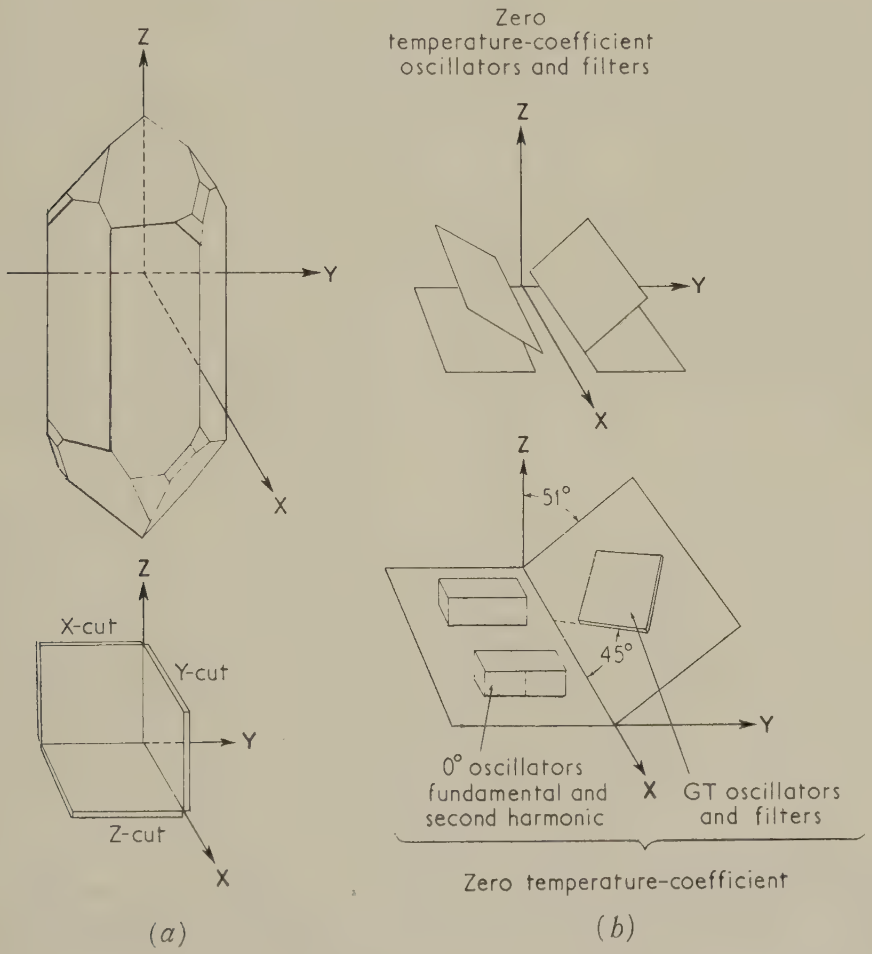


Fig. 3. Quartz crystal and principal cuts. (*a*) Crystal referred to orthorhombic axes, (*b*) Methods of obtaining crystals with zero temperature coefficients. (*Courtesy of W. P. Mason and Messrs. D. Van Nostrand Company Inc.*)

it is also important that the properties of the crystal do not change with temperature (e.g. crystals used to control oscillator frequency). In addition, suitable crystals must be readily available naturally or easily grown. The need to meet all or most of these requirements accounts for the use of such unfamiliar materials as ethylene diamine tartrate (EDT), for example, which is now used extensively in place of quartz. (Large crystals of natural quartz are now scarce and only limited success has been obtained in the growing of artificial quartz crystals.) As with quartz, the piezoelectric properties of EDT vary with direction and the crystals must be cut in particular directions with respect to their crystallographic axes, the actual directions depending upon the purpose for which the crystal is required (fig. 3).

Piezoelectric crystals are used as simple electromechanical transducers for the inter-conversion of acoustic and electrical energy. A bar or slab of quartz or other piezoelectric material can be made to oscillate by applying an alternating voltage between two electrodes attached to the ends of the bar. The resonant frequency of such a bar is determined only by the length, density and Young's modulus in the direction of oscillation, and with quartz and EDT there is so little internal damping that the Q factor is high, i.e. the resonance is very sharp. The amplitude of mechanical oscillation of the crystal therefore varies sharply with frequency near to the resonant frequency and so also does the amount of strain energy stored during each cycle. The dielectric constant of a quartz or EDT crystal, and the reactance of the condenser of which it forms part, therefore vary markedly with frequency (fig. 4). Thus the properties of the crystal can influence considerably the properties of the electrical circuit of which it forms part and piezoelectric crystals are extensively used for the accurate control of oscillator frequency and in narrow-band filter circuits. Quartz and EDT are used for these purposes because of their chemical stability and because they can be cut in such a way that their temperature coefficient of resonant frequency is almost zero.

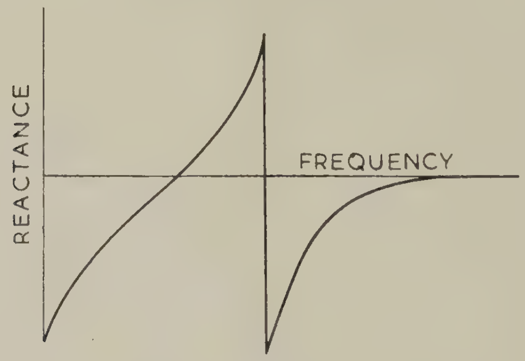


Fig. 4. The reactance of a piezoelectric crystal as a function of frequency. The discontinuity in the reactance corresponds to the resonant frequency of the crystal.

In other applications of piezoelectric crystals, e.g. for gramophone pick-ups, temperature stability is not so necessary and the most important requirement is a large electrical output for a given mechanical displacement or vice-versa. For devices that have to produce a large motion for a given voltage applied, Rochelle salt (sodium potassium tartrate tetrahydrate) was until recently the only useful material, but during the last war a new piezoelectric crystal—ammonium dihydrogen phosphate (ADP)—was developed. This material contains no water of crystallization and hence will not dehydrate, will withstand temperatures up to 100°c and will radiate considerable amounts of power without breaking down. These properties are so favourable that ADP has largely displaced Rochelle salt for many applications.

Although piezoelectricity is essentially a single-crystal property, a kind of artificial piezoelectricity can be developed in a particular class of polycrystalline materials, ferroelectric materials, so called because their properties are analogous with those of ferromagnetic materials. In piezoelectric materials the polarization is almost directly proportional to the first power of the applied field, and therefore changes sign when the direction of the field is reversed. There is, however, a component of the polarization that is proportional to the square of the applied field and this is found in all materials, whether or not they are piezoelectric. The corresponding change in dimensions is referred to as electrostriction : it is very small except in ferroelectric materials, and, being proportional to the square of the applied field, is not so generally applicable as the converse piezoelectric effect. In ferroelectric materials, however, the electrostrictive displacement can be made directly proportional to the applied field. This is

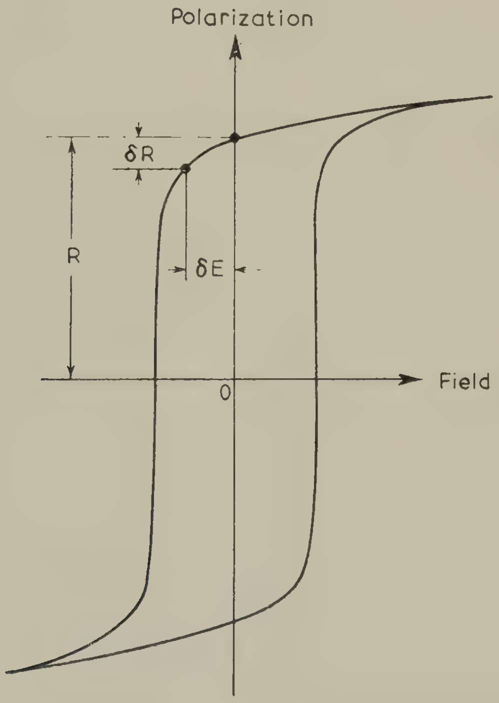


Fig. 5. Dielectric hysteresis loop for a ferroelectric material.

possible because ferroelectric materials exhibit hysteresis in the relation between polarization and field (fig. 5), and a remanent polarization remains even in polycrystalline material when the applied field is removed. If the effect of an applied field is to change the polarization from R to $R-\delta R$, say, and if the length of the material is at the same time changed from l to $l-\delta l$, then

$$\frac{\delta l}{l} \propto \frac{R^2-(R-\delta R)^2}{R^2} \propto \frac{2R\delta R}{R^2} \propto \frac{2\delta E}{R}$$

assuming that δR is approximately proportional to δE.

As with ferromagnetism, ferroelectricity is found over only a limited temperature range for any material. Rochelle salt, for example, is ferroelectric between −18°c and 24°c. The electrostrictive effect in Rochelle salt is about one-fifth of the magnitude of the converse piezoelectric effect for the applied voltages usually used and, in general, the electrostrictive effect is not of much practical interest in single crystals. But it is of interest in polycrystalline material (with which there is, of course, no piezoelectric effect) in which a remanent polarization has been produced by the momentary application of a large applied field, and a ceramic ferroelectric material, barium titanate, is being increasingly used in place of quartz or Rochelle salt for the generation of ultrasonic radiation, crystal microphones, gramophone pick-ups, etc. Its physical properties depend markedly upon temperature so that it cannot be used for frequency control or similar applications, but its large electrostrictive effect makes it very suitable for many other purposes. For example, a disc of barium titanate ceramic, loaded with a mass M, makes a simple accelerometer, a voltage being generated across the disc proportional to the force applied, i.e. to $M \times$ acceleration.

Barium titanate ceramic is also used as a dielectric material for condensers, the capacity of which is not required to be independent of temperature. As already explained, a large change in dimensions on the application of an electric field is associated with a large dielectric constant (as great as 1400 at ordinary temperatures for barium titanate) and high-capacity condensers of small physical size can be made.

Optical anisotropy

According to the electromagnetic theory of light there are, in an illuminated refracting medium, fluctuating electric and magnetic fields. The oscillating electric fields induce dipoles in the atoms and ions (the effect of the magnetic fields can usually be ignored in this connection) and the optical properties of solids are largely determined by the effect of these induced dipoles on the light wave. In a material of refractive index 2·0, for example, the combined effect of the induced dipoles reduces the phase velocity of the wave to one-half of its value in vacuum. The optical properties of solid materials are thus associated, as are their piezoelectric properties, with the polarization produced by applied electric fields, and therefore with the dielectric constant. The relation between dielectric constant and piezoelectricity has already been mentioned, and that between dielectric constant K and the phase velocity of light in the material v is well-known [$v=c/\sqrt{(\mu K)}$ where c is the velocity of light in vacuum and μ is the magnetic permeability of the material].

In general, the optical properties of crystals vary with direction in the crystal. When a ray of light enters an isotropic medium it is refracted according to Snell's law and a single refracted ray is formed. When light enters an anisotropic medium, however, there are in general two refracted ray paths. The medium is said to be doubly refracting or birefringent, and there are two values of the phase velocity of electromagnetic radiation in any given direction. However, for most anisotropic materials it is found that there are two directions for which there is only one phase velocity. These directions are called optic axes and the material is said to be biaxial. (The optical properties of crystals depend both on the nature of the constituent ions and on their mutual arrangement. The direction of the optic axes therefore depends upon the way in which the ions are arranged with respect to the principal axes of the crystal.) As a special case there are also uniaxial crystals for which there is only one optic axis and it is with uniaxial crystals that we shall be concerned in what follows. The practical applications of optical anisotropy are associated with the fact that, in directions other than the optic axis, only plane polarized waves with the direction of polarization in a particular orientation can be propagated : the directions of polarization in the two refracted rays are at right angles to each other.

Polarizing prisms (of which the Nicol prism is a well-known example) are usually made from calcite crystals so arranged that only one of the refracted rays can be transmitted so that plane polarized light is produced. In some crystals the two refracted rays are differently absorbed, a phenomenon known as dichroism. In tourmaline (a complex silicate), for example, one of the rays is very strongly absorbed and a single crystal plate cut parallel to the optic axis can be used to produce partially polarized light. Unfortunately the absorption of the unwanted ray is not complete but the same principle was employed, using orientated ultramicroscopic crystals of herepathite (iodosulphate of quinine) carried in a transparent film, in the early forms of Polaroid which has replaced polarizing prisms for many purposes. If a high degree of polarization is required, however, single crystal prisms must still be used.

In directions perpendicular to the optic axis the two refracted rays in a uniaxial crystal travel along the same direction but at different speeds. Thus if plane polarized light enters at normal incidence into a crystal cut parallel to the optic axis, the two rays will emerge from the plate with a phase difference between them and, since they are coherent and travelling along the same path, they are superposed to give plane, circular or elliptically polarized light depending upon the phase difference between them. As the thickness of the crystal plate is increased from zero the phase difference increases and for particular thicknesses of crystal a phase difference of $\pi/2$ can be introduced. Such a crystal is called a quarter-wave plate and will turn plane-polarized light into circularly-polarized light. Quarter-wave plates are often made of quartz or mica (a complex aluminium silicate) crystals. (Strictly mica is biaxial but can be obtained in a form that is almost uniaxial.) Half-wave, full-wave and other plates can also be made. Quartz wedges of variable thickness are used in the Babinet and Soleil compensators for the analysis of polarized light.

With some crystals lacking a centre of symmetry (e.g. quartz) it is found that if plane polarized light is incident normally on to a plate cut perpendicular to the optic axis, the emergent beam is also plane-polarized but its plane of polarization is not the same as that of the incident beam. The rotation of the plane of vibration is proportional to the thickness of the crystal plate and crystals that rotate the plane of polarization in this way are said to be optically active. The phenomenon is associated with a particular kind of lack of symmetry of the structure and optically-active crystals can usually exist in two forms, right-handed and left-handed. These are equivalent but not identical—the structure of one is the mirror image of the structure of the other (enantiomorphism). The two forms rotate the plane of polarization in opposite directions. (The optical activity of liquids or solutions is caused by the same kind of ' handedness ' in the molecules themselves and it is always much smaller than that of crystals in which it is the lack of symmetry of the arrangement of the molecules in the crystal that is primarily involved.)

2. Crystal perfection

In a perfect crystal each equivalent atom would have identical surroundings. However, even the best crystals contain many faults of various kinds, the principal ones being (*a*) vacant atomic sites, (*b*) impurity atoms, included either substitutionally or interstitially, (*c*) errors of atomic arrangement or dislocations. Single crystals can be grown only under conditions such that the number of such faults is relatively small and polycrystalline materials usually contain far more faults per unit volume than do single crystals. In addition the atomic arrangement in polycrystalline materials is interrupted by grain boundaries which also often contain impurities excluded from the growing crystals. The greater purity—chemical and physical—of single crystals enables them to be used for many purposes for which polycrystalline materials are unsuitable.

Optical properties

The dipoles induced by electromagnetic radiation in the material through which it is passing not only affect the nature of the radiation transmitted but can absorb energy from the incident wave. The nature and extent of the absorption depends upon the interaction between adjacent dipoles and thus upon the composition and structure of the material. For example, glasses that are transparent throughout the visible range show very strong absorption in the ultra-violet and infra-red. Furthermore, any disturbance of the crystal structure will, in general, modify the absorption characteristics. Thus many alkali halide crystals, which are colourless when pure, can be coloured by the addition of an excess of the metallic ions or of ions of the transition elements.

Glass absorbs infra-red and ultra-violet radiation, and prisms and lenses for use with these radiations must be made from other materials. The correct absorbing properties can be obtained only in crystalline materials and, in order to avoid unwanted absorption and internal scattering, these materials must be used in the form of pure single crystals. The optical properties of single crystals are of interest in the visible range too ; the brilliance of diamonds and other gem stones, although enhanced by the method of cutting, is associated also with the perfection and purity of the single crystals.

Prisms and lenses for use in ultra-violet spectroscopy are now made from single crystals of sodium chloride, potassium chloride, calcium fluoride or lithium fluoride, almost all being grown artificially ; quartz is also used but its optical anisotropy leads to complications. For infra-red spectroscopy, prisms and lenses made from crystals of sodium chloride, potassium bromide, potassium chloride, potassium iodide, silver chloride, thallium bromide-iodide, lithium fluoride or calcium fluoride are used.

It should be noted that all of the above-mentioned synthetic crystals are cubic and optically isotropic. It is not, therefore, necessary to have the path of the light parallel to any particular crystal direction. Indeed, rolled sheets of polycrystalline silver chloride are sometimes used as windows in infra-red apparatus.

Electrical properties

The electrical properties of semiconductors such as germanium or silicon depend markedly upon the nature and the amount of the impurities and also upon the degree of perfection of the atomic arrangement in these materials. In the pure state there are, at ordinary temperatures, very few conduction electrons available, but by the controlled addition of small amounts of the right kind of impurity, any desired resistivity may be obtained (fig. 6). But any imperfections in the atomic arrangement involve distortions of the electrostatic field inside the semiconductor which can have the effect of 'trapping' a conduction electron or of allowing it to combine with an ionized atom so that it is no longer available

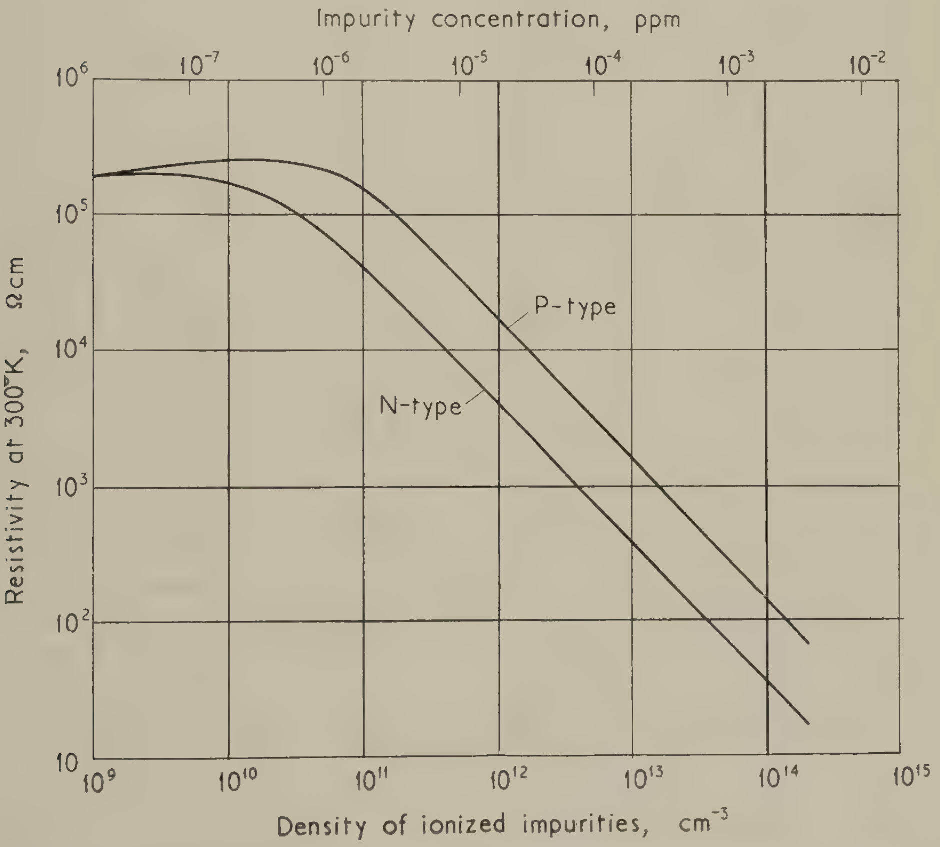


Fig. 6. The electrical resistivity of silicon containing different amounts of impurity. *(Courtesy of J. M. Wilson and Messrs. Heywood & Co. Ltd.)*

to contribute to the conductivity of the material. The performance of devices such as crystal diodes and transistors is very dependent upon free movement of the current carriers and such devices must be made from germanium or silicon which is as free as possible from all types of structural defect. In order to obtain the necessary purity and structural perfection, these materials are prepared as large single crystals grown under carefully controlled conditions.

But whilst for these applications lattice defects are unwanted, crystals are also used in applications in which the creation of lattice defects within the crystal is used as a means of detecting the presence and energy of ionizing particles or radiation.

In an ionization chamber the ionization created by the absorption of a particle or quantum of radiation within the chamber is measured in terms of the quantity of electricity that is carried across the chamber as the ions produced move under the influence of the applied voltage. An insulating crystal can be used instead of an ionization chamber, the principal differences being that the stopping power of the crystal is appreciably greater than that of a gas-filled counter (so that more energy can be absorbed from particles of high energy), and that the time response of the crystal counter is less than that of an ionization chamber and hence particles can be counted at a greater rate. Crystals of diamond have been used as crystal counters, but the obvious difficulty of obtaining crystals of large size limits the usefulness of this type of counter and a more popular method of detecting radiations is the scintillation counter.

When a photon is absorbed by an atom, the atom is raised to an excited state : this state is usually not stable and the atom will return to its original state giving out in the process, as a photon of radiation, the energy absorbed. The frequency of the radiation emitted is always lower than that of the exciting radiation. Thus a high-energy particle traversing a crystal may excite an atom and the energy absorbed may be subsequently re-emitted as visible radiation which can be detected by a photoelectric cell. Such an arrangement of absorber and detector forms a scintillation counter. A large transparent crystal enables a large absorbing volume to be obtained and a photomultiplier tube is used to detect the flashes or scintillations. Crystals of anthracene, napthalene, stilbene, terphenyl or alkali halides are grown for use in scintillation counters, the requirements being a large crystal, transparent to the radiation emitted (and therefore free from flaws or grain boundaries), and emitting radiation of a suitable frequency. This latter may often be controlled by the addition of suitable impurities. This is referred to as activation and, for example, potassium iodide crystals activated with thallium iodide are used.

Masers

Single crystals of substances such as ruby or emerald (forms of alumina containing small amounts of chromium as impurity), potassium chromicyanide and some others have recently been used in the high-frequency amplifying devices known as masers (Microwave Amplification by the Stimulated Emission of Radiation). These are discussed in the article by K. W. H. Stevens in the October 1960 issue of this journal. A more recent development (in the Bell Telephone Laboratories) is the optical maser which works on the same principle and uses a crystal of ruby to produce a very intense, very monochromatic, coherent and extremely narrow beam of light.

Reference has already been made to the effect of impurity atoms on the optical properties of crystals. Closely associated with this is the phenomenon of fluorescence : a fluorescent material is one which absorbs light of a particular wavelength and then, later, spontaneously emits light of a longer wavelength. For example, when a ruby is illuminated with green light, it fluoresces and emits red light. The green light excites the impurity (chromium) atoms to an energy state higher than that in which they normally exist : the atoms rapidly relax to an intermediate, metastable state and then, more slowly, to their original state. In making this second energy jump each atom emits a quantum of radiation of a wavelength corresponding to the difference in energy between the intermediate and final states of the atom. If, however, after having been excited, the atoms are irradiated with light of the frequency that they will later emit, they can be stimulated to make the second energy jump much more readily.

A single crystal rod of synthetic ruby forms the heart of the optical maser. The ends of the rod are polished until they are extremely flat and parallel, and they are then coated with a reflecting layer of silver thin enough to be slightly transparent. The rod is illuminated by intense flashes of white light, the green component of which excites the chromium atoms to their higher energy level. They fall back rapidly to the intermediate level, but spontaneous emission from this level is relatively slow. However, the light emitted by the few spontaneously emitting atoms is of the proper frequency to stimulate emission by the remainder of the excited atoms. The silvered ends of the ruby trap some of this light, which is then reflected back and forth so that it has ample opportunity to stimulate emission from the other atoms. As a result of this stimulation, intense flashes of red light—each lasting about one microsecond—are produced, the flashes continuing as long as the crystal is illuminated by the white light.

Only those waves travelling along the axis of the rod are amplified by the stimulated emission : all the other waves miss the end plates and thus spend little time in the crystal. Since the end plates are small in comparison with the distance between them, only the waves travelling accurately along the axis are amplified. For this reason the output beam that leaks through the partially transparent ends of the crystal has almost all of its intensity within a very narrow cone. Within this narrow cone and frequency band, the beam of light emitted by the ruby optical maser is more than a million times brighter than the sun. Two other features of the light emitted are of special interest. It is much more nearly monochromatic than the usual fluorescence of ruby. This is a direct result of the stimulated emission ; the closer the stimulating radiation is to the correct wavelength, the more effective is the stimulation. Thus almost all the radiation emitted is of very nearly the same wavelength. The light from the maser is also coherent, i.e. the ' crests ' of the waves from all the emitting atoms are exactly in phase in the light emitted. This, too, is a consequence of the stimulated emission. These features of the light emitted by the optical maser are expected to make it useful for communication purposes since it is now possible, in principle, to extend our present communication technology to optical frequencies.

READING LIST

Some suggestions for further reading :

The Crystalline State, ed. by W. H. Bragg and W. L. Bragg. Vol. I: A general survey, by W. L. Bragg (London : G. Bell & Sons Ltd., 1933).

Plasticity of Crystals, by E. Schmid and W. Boas (London: F. A. Hughes & Co. Ltd. 1950).
An Introduction to Solid-state Physics, by C. Kittel (New York: John Wiley & Sons, Inc.; London: Chapman & Hall Ltd., 1953).
Piezoelectric Crystals and their Application to Ultrasonics, by W. P. Mason (New York: D. Van Nostrand Co. Inc., 1950).
Electrons and Holes in Semiconductors, by W. Shockley (New York: D. Van Nostrand Co. Inc., 1950).
A Textbook on Crystal Physics, by W. A. Wooster (Cambridge: University Press, 1938).

The Author:

Mr. A. E. De Barr is Director of Research of the Machine Tool Industry Research Association. He has published papers on x-ray crystallography, magnetic materials, the physics of textile processes and other branches of applied physics. His interests include technical education in which he takes an active part.

Neutrino and Antineutrino Losses in Nuclear Processes

by G. M. LEWIS
Department of Natural Philosophy, The University, Glasgow

SUMMARY

Energy, momentum and angular momentum losses in β-decay processes are ascribed to neutrinos and antineutrinos. These penetrating neutral particles, with peculiar symmetry properties, are responsible for sizeable irrecoverable energy losses in nuclear reactors and thermonuclear processes.

1. INTRODUCTION

About a thousand unstable nuclei are now known which undergo some form of β-decay. These maladjusted nuclei can be regarded as having an excess of neutrons, or an excess of protons, over the stable nuclei into which they ultimately transform. Typical of the process is the decay of the free neutron which is observed to decay into a proton and an electron. In this transition, of half life ~12 min, the electron can have any energy value between zero and 780 KEV ($\sim 10^{-6}$ ergs). The missing part of the energy in such a β-decay process is associated with the emission of a neutral particle. This particle is generally referred to as the neutrino (ν) in β^+ (positron) decay, and as the antineutrino ($\bar{\nu}$) in β^- (electron) decay. Thus the conversion of a proton in a nucleus into a neutron in the nucleus may be expressed:

$$p \rightarrow n + \beta^+ + \nu. \qquad (1)$$

The conversion of a neutron in a nucleus into a proton in the nucleus may be written:

$$n \rightarrow p + \beta^- + \bar{\nu}. \qquad (2)$$

As discussed below, experiments now testify to the differences between the neutrino and the antineutrino; the neutrino spins left handedly about its line of flight, and the antineutrino spins right handedly about its line of flight; i.e. they have opposite helicities (fig. 1).

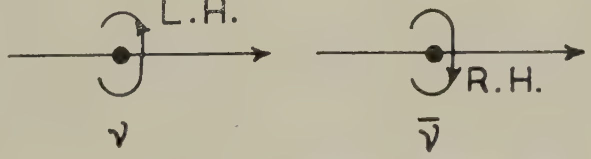


Fig. 1. The neutrino and antineutrino are particles of opposite type. The neutrino (ν), emitted in β^+ decay, spins left handedly about its line of flight. The antineutrino ($\bar{\nu}$), emitted in β^- decay, spins right handedly about its line of flight.

The maximum β-energy accords closely with that deduced from mass difference methods on the nuclei concerned. These and other findings indicate that the rest mass of the neutrino or antineutrino is less than 1/1000 that of the electron. This suggests their rest masses are zero and that they move with the speed of light.

Analogous to the decay process (1), and also responsible for proton-to-neutron conversion in the nucleus is the electron capture process where an atomic orbital electron, usually a K-shell electron, is captured by the proton of the nucleus, according to:

$$p + \beta^- \rightarrow n + \nu \tag{1'}$$

e.g.

$$A^{37} + \beta^- \rightarrow Cl^{37} + \nu.$$

Equation 1′ is consistent with adding the opposite or antiparticle β^- of the positron, to both sides of eqn. (1).) Here the decay energy is divided, unequally, between two bodies only, the neutrino which takes most of it, and the massive nucleus which shows a small unique recoil. This recoil, observed experimentally, serves as a check on the postulated existence of the neutrino. For β-decay in general (cf. relations 1 and 2), the recoil energy is of course not unique, but recoil measurements have yielded information on how the directions of emission of the β-particle and antineutrino are related (fig. 2).

Finally, apart from their rôle in the β-decay of the nucleons, neutrinos and antineutrinos are emitted in meson decay processes.

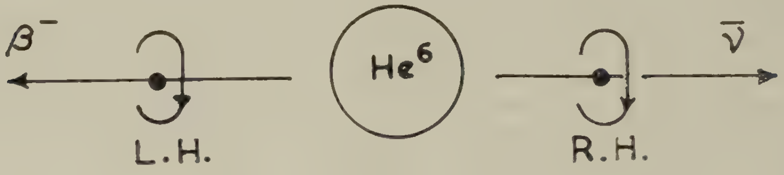


Fig. 2. In the decay $He^6 \rightarrow Li^6 + \beta^- + \bar{\nu}$, a fast β^- ray and $\bar{\nu}$ are emitted predominantly in opposite directions. The nucleus suffers a spin 1 change in the transition. (c.f. Herrmannsfeldt *et al.*) The emitted particles take up the spins shown.

2. SPIN AND MOMENTUM PROPERTIES OF THE ν AND $\bar{\nu}$ PARTICLES

Before 1956, attempts to specify a distinction between neutrinos and antineutrinos were partly based on double β-decay work and no definite conclusions could be reached by that method at the time.

A new approach to the subject was provided by the revolutionary ideas of Lee and Yang, that the helicities of the particles ν and $\bar{\nu}$ were irrevocably opposite (fig. 1). Such a suggestion implied that the β-rays from the source would also tend to possess a definite spin about their line of flight. This can be seen as follows.

Recoil measurements of the decay of gaseous He^6 by Herrmannsfeldt and others, show that a β^- particle and the antineutrino $\bar{\nu}$ tend to be emitted predominantly in opposite directions as shown in fig. 2, particularly if the β^- ray is fast. This particular nucleus suffers an angular momentum change of 1 (in units of $\hbar$) in the decay. The fast β^- particle which has a spin $\frac{1}{2}$, and the $\bar{\nu}$ particle which must also have a spin $\frac{1}{2}$, must then spin together so as to provide the spin 1 change. Because they are travelling in opposite directions their helicities are opposite, i.e. if the antineutrino were right handed about its direction of motion, the fast β^- particle would be expected to be left handed about its direction of motion. Similar conclusions will hold for other β^- emitters with spin change 1, e.g. Co^{60}.

The helicity of such β^- rays has been determined by a number of workers, for instance by scattering the β^- rays by the partially aligned electrons of magnetized iron. Such fast β-rays, spin left handedly about their direction of travel, indicating that the antineutrinos have right handed helicity. Similar experiments with β^+ emitters, where the nucleus had a spin 1 change showed opposite results, the fast positron had right handed helicity and the neutrino had left handed helicity. It seems likely, further, that fig. 1 is a valid description of the ν and $\bar{\nu}$ particles in all β-decay processes.

3. NEUTRINO ENERGY FLUXES IN REACTOR PROCESSES

In highly energetic β-decay processes, where the rest energy of the electron is not a significant quantity, the average energies removed by the β-rays and neutrinos (or antineutrinos) are alike. For less energetic β-processes, as much as 0·75 of the disintegration energy is, on average, lost to these neutral particles. Moreover for an electron capture process it can be much greater, e.g. for K capture in C^{11} all but 400 ev of the 2 MEV disintegration is carried away by each neutrino.

In a reactor the basic constituent uranium is a nucleus overburdened with neutrons. The fission fragments in turn tend also to contain excess neutrons. These fragments decay therefore with the emission of antineutrinos. Any free neutrons which escape afford a further source of antineutrinos.

The antineutrino flux in nuclear reactors based on the U^{235} thermal neutral fission process has been assessed by Muelhause and Oleksa. Though the main energy release occurs in the fission of this nucleus (~200 MEV), further energy is given out in the five or so successive β^- decays of the various fission fragments (fig. 3). In this way some 10 MEV of energy is communicated to anti-

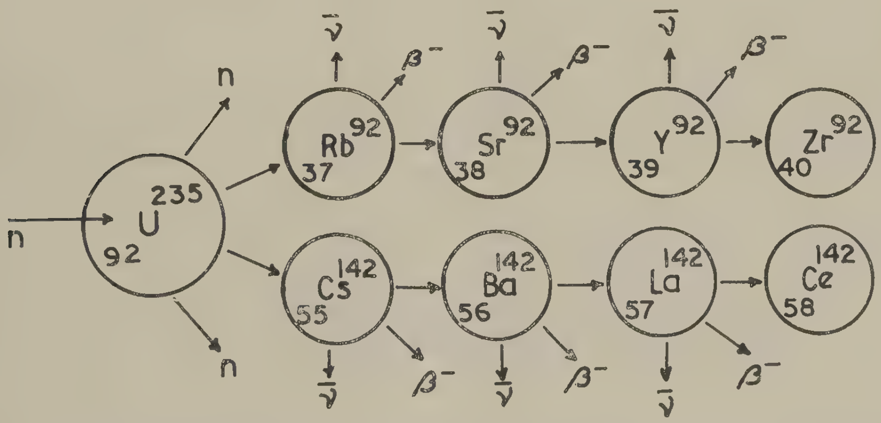


Fig. 3. One of several possible fission processes, in which the excess neutrons of the two fission fragments are successively transformed, emitting antineutrinos.

neutrinos per fission, and about 5 per cent of the reactor energy is therefore lost to them. In this way it is estimated that about 2×10^{11} antineutrinos per sec are emitted per watt generated in the pile.

The large flux obtainable from a pile prompted a successful attempt by Cowan and Reines to intercept antineutrinos in flight. Among the most likely of collision processes is that of inverse β-decay, the inverse of relation (2) viz:

$$p + \beta^- + \bar{\nu} \rightarrow n.$$

This can be re-expressed

$$p + \bar{\nu} \rightarrow n + \beta^+.$$

A high antineutrino flux exceeding 10^{13} antineutrinos/cm² × sec from the Savannah River thermal neutron chain reactor was used. A very few antineutrinos were absorbed by the protons of a large water target. The neutrons and the annihilation radiation from the positrons formed were detected by large tanks of scintillating fluid, viewed by several hundred photomultipliers. The measured cross section for the interception has been recently given as $(1{\cdot}1 \pm 0{\cdot}4) \times 10^{-43}$ cm², and is near to the expected value. This implies that an antineutrino in a beam has a chance of only about 10^{-20} of being picked up by 1 cm thickness of liquid hydrogen in its path. Such a minute chance indicates that antineutrinos (and neutrinos) could traverse the matter of many millions of stars without being intercepted.

It should be noted in conclusion that reactor antineutrinos could not be expected to take part in inverse reactions of type:

$$\nu + \mathrm{Cl}^{37} \rightarrow \mathrm{A}^{37} + \beta^- \quad \text{(cf. eqn. (1′))}$$

in accord with experiments by Davis.

4. THERMONUCLEAR LOSSES, AND WIDER PROBLEMS

One of the main stellar processes suggested is the hydrogen chain reaction initiated by the collision process:

$$\mathrm{H}^1 + \mathrm{H}^1 \rightarrow \mathrm{H}^2 + \beta^+ + \nu.$$

This is essentially the β^+ decay of an unobserved He². The process is followed by further additive reactions leading to formation of helium. The overall changes, and energy produced, may be expressed:

$$4\mathrm{H}^1 \rightarrow \mathrm{He}^4 + 2\beta^+ + 2\,\nu + 26 \text{ MeV}.$$

An identical overall change and energy release occurs in the carbon cycle. This starts by the process

$$\mathrm{C}^{12} + \mathrm{H}^1 \rightarrow \mathrm{N}^{13}(+\gamma \text{ radiation}) \rightarrow \mathrm{C}^{13} + \beta^+ + \nu$$

and ends after four successive hydrogen captures with the reaction:

$$\mathrm{N}^{15} + \mathrm{H}^1 \rightarrow \mathrm{C}^{12} + \mathrm{He}^4.$$

Here again hydrogen is converted into helium with the release of 26 MEV. But in this latter process different elements take part in the relevant β^+ decay processes. The neutrino losses in the two methods are somewhat different. In the first (H–H) chain reaction this loss is 0·5 MEV, and in the carbon cycle it is about 2 MEV.

These and other fusion processes are responsible for the enormous flux of neutrinos from the sun at the earth's surface, day and night, of order 10^{11}/cm² × sec.

Finally, perhaps the most interesting process in the present context is that proposed by Gamow and Schoenberg for stellar explosions (supernovae) in which the energy of a star is rapidly dissipated into a vast escaping flux of neutrinos and antineutrinos (fig. 4). Thus a highly energetic electron could be captured by a proton (according to relation 1′) even in a stable nucleus (cf. O^{16}), producing a neutrino (ν) which escapes and an unstable nucleus (e.g. N^{16}). This unstable nucleus (N^{16}) then decays back to the stable nucleus (O^{16}) and an antineutrino which also escapes, and so on repetitively. In a star this could give rise to violent energy emission.

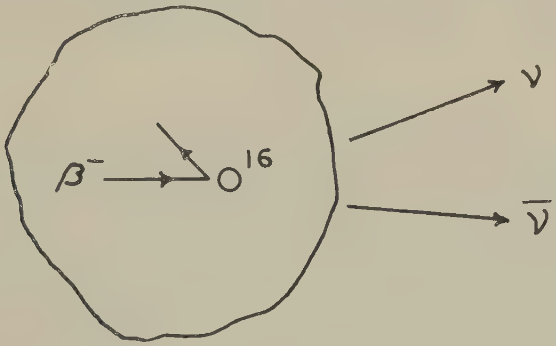


Fig. 4. Representing an event in the depths of a star, and the ν and $\bar{\nu}$ escaping unimpaired from it. Neutrino, or antineutrino losses of order a few per cent can arise in several chain processes. Gamow and Schoenberg have however indicated that stellar energy conversion into ν, $\bar{\nu}$ energy fluxes of catastrophic magnitude can occur. Such a process is exemplified by the capture of an energetic electron by stable O^{16}, viz: $\beta^- + O^{16} \rightarrow N^{16} + \nu$, followed by the 7-second decay: $N^{16} \rightarrow O^{16} + \beta^- + \bar{\nu}$; repetitively.

In conclusion one is left to contemplate the ultimate fate in the universe of the vast streams of these elusive neutral particles, which are emitted in so many nuclear processes, and which pass relatively undisturbed through matter of celestial dimensions.

REFERENCES

ALLEN, J. S., 1958, *The Neutrino* (Princeton University Press).

GAMOW, G., and CRITCHFIELD, C. L., 1949, *Theory of Atomic Nucleus and Nuclear Energy Sources* (Oxford: Clarendon Press).

Conference on Weak Interactions, 1959, *Rev. mod. Phys.*, **31**, 782–838.

Radio Emission from Interstellar Neutral Hydrogen

by R. D. DAVIES

Jodrell Bank Experimental Station, University of Manchester

SUMMARY

The emission line for neutral hydrogen at 1420 Mc/s is the only line so far discovered in radioastronomy. Since its mechanism of origin is completely understood, observations of this line provide direct information about conditions in interstellar space such as temperatures, densities and velocities. Extensive investigations of our own Milky Way system have shown clearly that it is a spiral galaxy similar to, but rather smaller than, the great spiral nebula in Andromeda. Our knowledge of the spiral structure of galaxies is far from complete ; hydrogen-line measures of high-speed expansions in the centre of the Milky Way system may provide a clue to the understanding of this problem. In addition, determinations of the hydrogen content of different types of galaxy reinforce current theories which suggest that irregular galaxies evolve through the spiral form to elliptical galaxies.

1. INTRODUCTION

Hydrogen is the chief constituent of the stars which make up the Milky Way, the galaxy to which our Sun belongs. The fusion of this hydrogen into helium provides the source of the large amounts of energy emitted from stars. However, before this hydrogen condenses gravitationally into stars it lies in interstellar space in its invisible neutral form. The existence of neutral hydrogen having a density of about one atom/cm^3 was inferred from the known regions of ionized hydrogen which surround very hot young stars of types O and B. Such regions arise from the ionization of neutral hydrogen by ultraviolet starlight and they become visible when protons and electrons combine to form an excited hydrogen atom which falls to a lower energy state giving the first Balmer line (Hα) of hydrogen which is in the red end of the visible spectrum. On the other hand the bulk of the hydrogen in the interstellar medium is in the neutral ground state and has no emission in the visible range.

A new interest in the interstellar medium was aroused by the pioneer measurements of Jansky, Reber and others during the 1930's and the early 1940's of the radio waves emitted from the Milky Way. These were supposed rightly to come from the regions between the stars although the mechanism of emission was a mystery. These discoveries stimulated van de Hulst at Leiden in 1944 to consider whether there might be emission or absorption lines in the radio domain which would prove as useful as the lines in stellar spectra are to the optical astronomer.

The most promising radio line appeared to be that from neutral hydrogen at a frequency of 1420·406 Mc/s and a wavelength of 21·1 cm. It was considered that neutral hydrogen at the estimated density of 1 atom/cm^3 should be detectable.

However it was not until 1951 during the period of intense growth of radioastronomy after the War that van de Hulst's predictions were confirmed. Three independent research groups at Leiden, Harvard and Sydney reported their discovery of neutral hydrogen emission within a few months of one another. What was to become a fruitful new branch of radioastronomy had burst into growth.

2. The mechanism of emission

The radio emission line from neutral hydrogen comes from atoms in their lowest energy state, the ground state denoted by $^2S_{\frac{1}{2}}$. This state is divided into two very closely spaced levels called hyperfine levels which arise because the atom has different energies if the nucleus spins on its axis in the same or the opposite sense to the electron. In equilibrium conditions, which are those we will be dealing with in this article, for every one neutral hydrogen atom in the lower energy level there will be three in the higher energy level.

Any hydrogen atom in the upper of these two levels will only fall back spontaneously into the lower level after the elapse of eleven million years on the average. This energy change is consequently known as a metastable transition. Collisions between neutral hydrogen atoms keep the levels populated in the ratio 3 : 1 since there are many collisions within this long transition time. For example a given neutral hydrogen atom in interstellar space will suffer a collision once every fifty years or so. But since only one collision in eight re-excites an atom from the lower to the higher state this ' low energy ' atom must wait on the average four hundred years before being transferred to the high energy state. After a similar period it will return to the lower state as a result of collisions. This history will continue for some eleven million years before the atom eventually falls spontaneously (that is independently of collisions) from the higher to the lower state to give a photon of 21 cm wavelength.

As the density of interstellar neutral hydrogen is only of the order of one/cm^3, there will only be one photon per eleven million years/cm^3. Such a slow rate of energy release becomes detectable in practice because of the immense size of our Galaxy. In the direction of the galactic centre in Sagittarius, for example, neutral hydrogen atoms extend for 6×10^{22} cm.

The amount of 21 cm radiation from a given cloud of neutral hydrogen atoms is measured in terms of a quantity called the optical depth which is a function of the density in the line of sight and the temperature within the cloud. If the optical depth is large (say greater than unity) radiation from behind the cloud will not pass through it and the cloud will look like a radiator whose surface temperature is the kinetic temperature of the neutral hydrogen. In radioastronomy the emission from a given region of space at a given wavelength is most often given as a brightness temperature, which is the temperature of the black body radiator which radiates the same amount of energy at that wavelength as the given region. The brightness temperature of a neutral hydrogen cloud will only equal its kinetic temperature if it has large optical depth. The observed kinetic temperature of interstellar neutral hydrogen estimated from the regions of greatest brightness temperature is rather greater than 100°K.

3. TECHNIQUES OF HYDROGEN LINE OBSERVATION

The signals received from extra-terrestrial neutral hydrogen are extremely weak, giving brightness temperatures of 100°K at the maximum and perhaps 2°K for many investigations. Due to the inherent slight inefficiency of a parabola as a collecting aerial, the aerial temperature of a region is about half its brightness temperature. The aerial temperature is said to increase by 1°K due to the presence of an emitting region in the aerial beam if the increase of the signal passed on to the receiver is the same as the increase which results from heating the aerial by 1°K. Thus the requirement for much present-day hydrogen line work is a receiver which will distinguish input signals of 1°K in the presence of 1000°K of noise introduced in the first stages of the receiver.

One technique used in radioastronomy for measuring small signals involves switching the receiver between the aerial and a resistor at room temperature and subtracting the two signals received ; this eliminates the large 'set noise' deflection from the record and increases the stability of the set. A similar technique is used in neutral hydrogen studies where the receiver is switched between two narrow band channels, one which contains the neutral hydrogen signal and one which lies outside the frequency of neutral hydrogen emission (see fig. 1*a*). The difference signal then refers to neutral hydrogen emission and its spectrum may be obtained by scanning the channels in frequency. Such a spectrometer record taken with a 4 km/sec bandwidth at the new galactic coordinates* $1''=84°$, $6''=0^0$ as shown in fig. 1*b*. The hydrogen which is traceable over a velocity range of 120 km/sec is grouped at three distinct

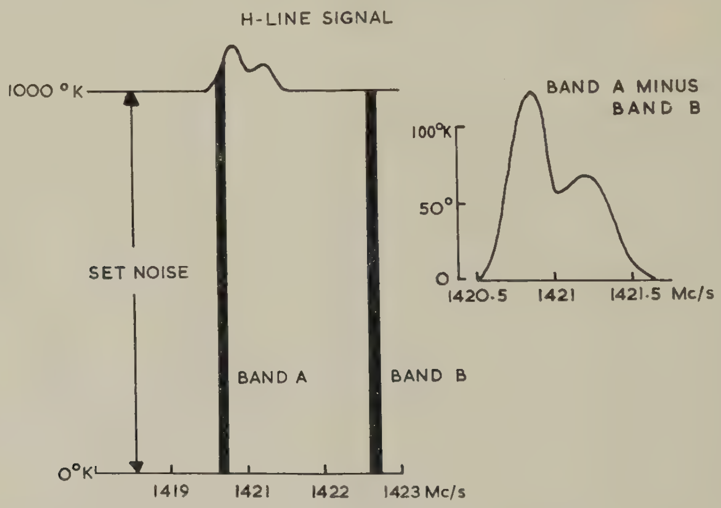


Fig. 1*a*. The signal from interstellar hydrogen is much less than the noise introduced by the set itself. A hydrogen line receiver gains its stability by subtracting the set noise at a frequency outside the hydrogen band from the noise received at the hydrogen line frequency. The resultant difference profile is shown enlarged.

---

* The new system of galactic coordinates is chosen such that the galactic centre is at longitude (l'') zero and latitude (b'') zero. In the old system (l', b') the centre is at $l'=328°$, $b'=1\cdot5°$.

velocities which correspond to gas concentrations at three different distances from the observer. The conversion from frequency to a distance will be explained in the next section.

It will be noticed that the spectrum has a ripple upon it which is due to the noise fluctuations inherent in the receiver. This noise ripple will have to be decreased if a large improvement in sensitivity is to be achieved. Indeed during the last few years important new equipmental developments have occurred which make it feasible to detect aerial temperatures of 1/10°K and less. These devices called masers and parametric amplifiers have input noise temperatures of 100°K or less as compared with 1000°K for conventional receivers in the 1400 Mc/cs range.

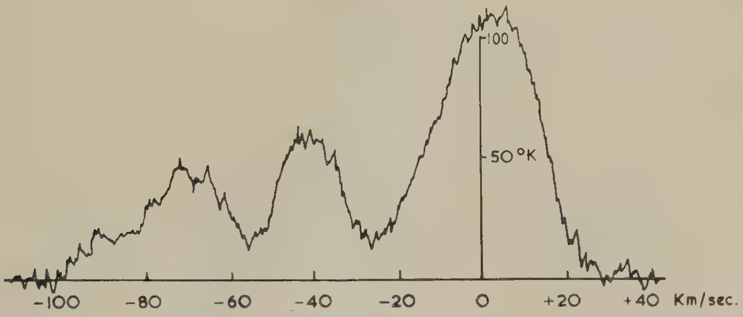


Fig. 1*b*. A profile observed at $l'' = 84°$ and $b'' = 0$ with a 15 kc/s bandwidth.

Another factor which sets a limit on any radio-astronomical observation is the angular resolution of receiving aerials which for 21 cm studies are normally parabolic reflectors with a dipole or horn feed at the focus. The bulk of the hydrogen line observations so far have used beamwidths of about 1 degree. New aerials have been completed or are in course of construction which have much smaller beamwidths. The 250 ft steerable radio telescope at Jodrell Bank has a beamwidth of 15 min of arc at 1420 Mc/s. This increased resolution in angle allows much interesting and important detail to be seen in features of our Milky Way system and also in external galaxies.

4. THE SPIRAL STRUCTURE OF THE GALAXY

Shapley in 1928 was the first to recognize that the Milky Way system was a disk-shaped assemblage of stars 50 000 light years† across. Although it was realized that the Milky Way was in many ways similar to the great spiral galaxies photographed with the large telescopes, its spiral structure could only be hinted at from the optical observations. The limiting factor which confines the optical astronomer to the nearer parts of the galactic disk (up to 5000 light years distant) is the obscuration produced by interstellar dust. It produces no such attenuation of radio waves.

Nevertheless optical measurements, including studies of star groups above the galactic plane, give a clear picture of the way in which the Galaxy rotates. The inner stars move as a solid disk ; the outermost stars move in Keplerian orbits as if they were independently moving about a point central mass, and giving the result that the most distant stars take the longest time to make a circuit of the galactic centre. This optical rotation curve is basic to any interpretation of the neutral hydrogen spectra in terms of the spatial distribution of gas in the Milky Way. The optical curve is shown schematically in fig. 2. The distances moved in 20 million years by stars at different distances from the galactic centre are indicated by the length of the arrows.

---

† 1 parsec = 3·26 light years = $3{\cdot}09 \times 10^{18}$ cm.; distances are given here in kiloparsecs (kpc).

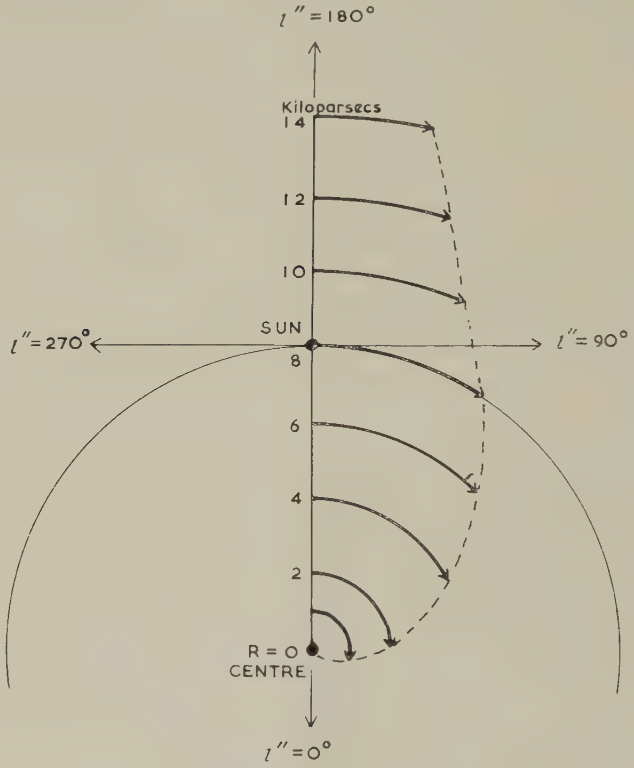


Fig. 2. A diagram of the Galaxy showing the distance moved by stars and gas at different distances from the galactic centre during a time interval of 20 million years.

The spiral nature of the Milky Way was traceable out to parts unseen by optical telescopes in H-line spectra taken at all points around the galactic plane by the Leiden group in the northern hemisphere and the Sydney group in the southern hemisphere. Figure 3 shows spectra obtained in the galactic plane at $l''=0°$, 90°, 180° and 270°. Concentrations of neutral hydrogen occur at different frequencies displaced relative to the frequency of stationary neutral hydrogen due to their movement in the line of sight. These frequency shifts may be converted to velocities by multiplying by $-1/4{\cdot}73$ km/s per kc/s ; a positive velocity denotes recession in astronomical nomenclature. These velocities for a given direction may in turn be converted into distances using the galactic rotation curve given above. When spectra for all directions are treated in this way a picture emerges of the distribution of neutral hydrogen throughout the Galaxy. Figure 4 is the map obtained by Oort, Kerr and

Westerhout. It will be noted that there are gaps in the distribution in the direction of the centre and the anticentre where the motion of all the hydrogen is perpendicular to the line of sight and consequently there can be no separation in distance.

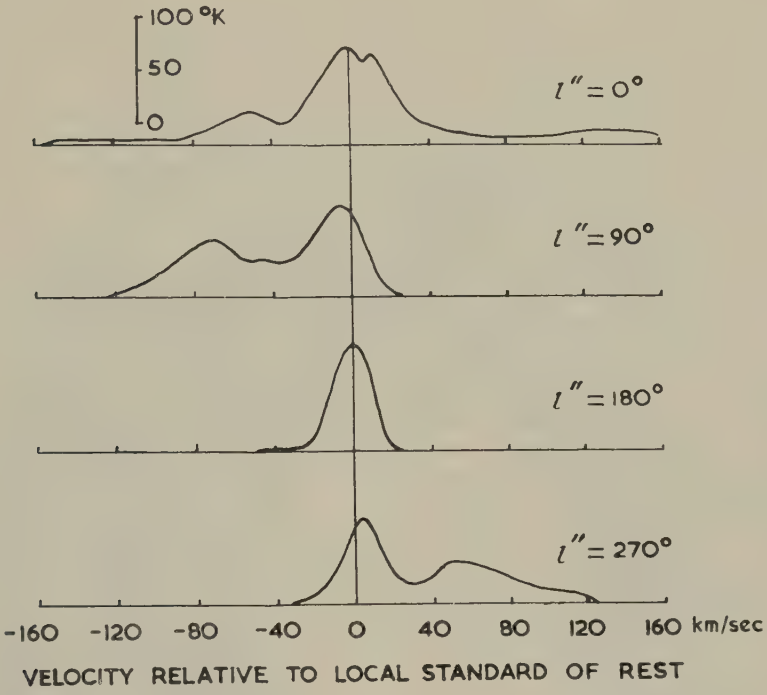


Fig. 3. Hydrogen line spectra observed at $l'' = 0°$, 90°, 180° and 270°. These directions are indicated in Fig. 2.

Several arm-like features are clearly marked in fig. 4 and perhaps the most conspicuous is the concentration of hydrogen known as the Orion arm which passes near the Sun and extends around the northern sky to the far side of the Galaxy. This may possibly be connected with the Carina arm in the southern sky which lies at a distance (R) from the centre of 8 kpc. The two northern arms lying outside the Orion arm are known as the Perseus and Outer arms. The northern and southern halves of the Galaxy are represented in different detail ; the southern observers prefer to show the general trend of the distribution rather than assigning a particular distance to every velocity detail shown in the spectra.

The inner parts of the Galaxy have to be mapped by a modification of the method employed above since the line of sight cuts a given constant R circle at two points, each giving the same radial velocity. This ambiguity can in most cases be satisfactorily resolved by examining the distribution of hydrogen in latitude ; the broad distribution is supposed to be nearby hydrogen and the narrow distribution is more distant hydrogen. A clearly defined arm is found at $R = 6$ kpc and is known as the Sagittarius arm. Other elongated features are also visible.

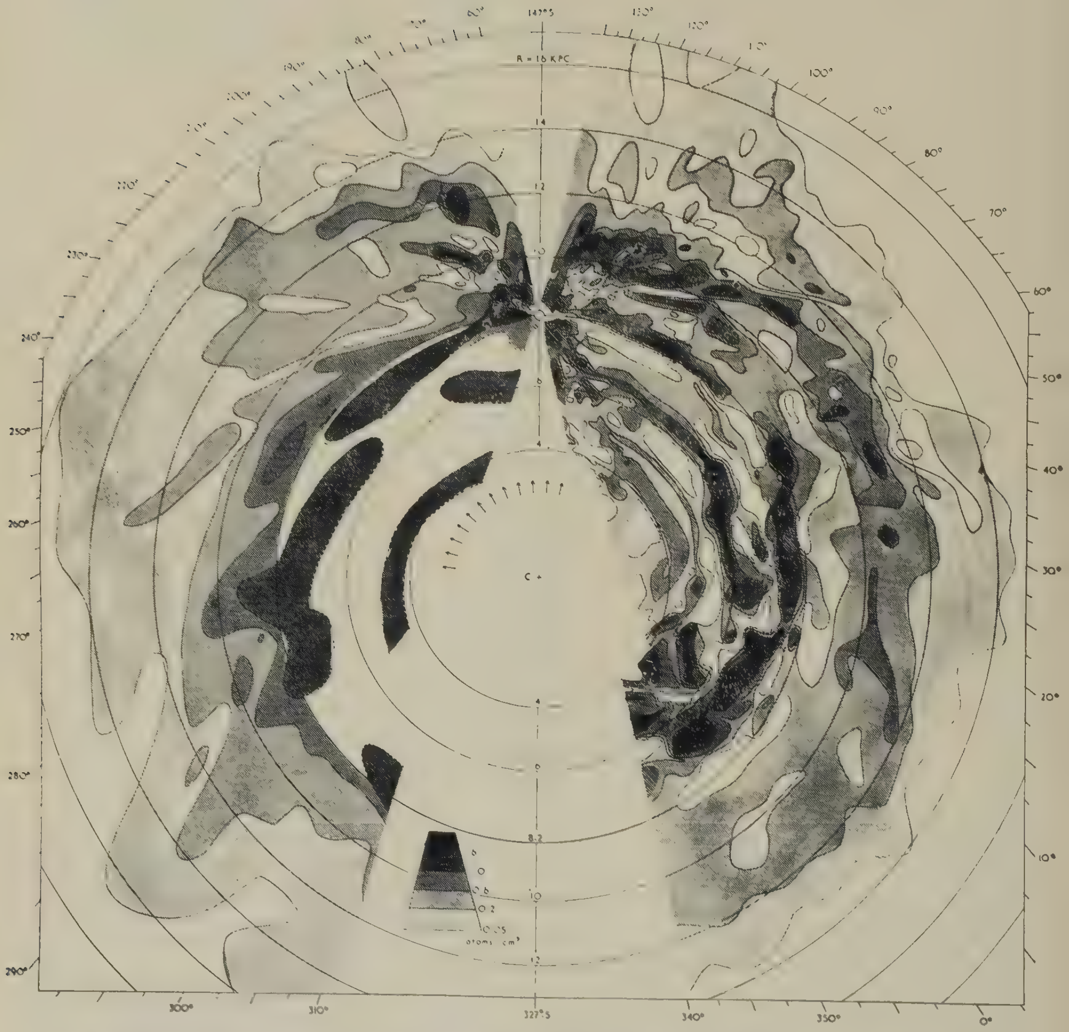


Fig. 4. The distribution of neutral hydrogen in the plane of the Milky Way system. The old galactic coordinate system (l', b') is used. (By courtesy of J. H. Oort, F. J. Kerr and G. Westerhout and the editors of *M.N.* of the Royal Astronomical Society).

It is of immediate interest to compare the neutral hydrogen spiral structure with any optical features of our Galaxy which may be expected to show spiral structure. Studies of external galaxies indicate that young bright stars of types O and B, ionized hydrogen regions and Cepheid variables are the most suitable. The distribution of these objects to distances of 4 kpc from the Sun is well established in some directions. Arm structure has been found, but it does not coincide well with the position of the neutral hydrogen arms. The optical arms in the solar vicinity are apparently inclined at about 20° to the hydrogen arms and moreover many O and B stars lie between the hydrogen arms. These discrepancies may not be real since there are errors in both the radio and optical distances. Optical distance determinations depend upon only a fair knowledge of the amount and size of the absorbing dust while the radio determinations depend upon the unproven assumption that the rotation curve of the galaxy is circularly symmetrical.

The overall impression of the Galaxy given by hydrogen-line measurements is of a thin disk only 0·220 kpc in depth but 25 kpc in diameter. In the central regions of the Galaxy where $R < 8$ kpc this disk is remarkably flat, but beyond $R = 8$ kpc there is an intriguing systematic tilting of the hydrogen above the plane in the northern hemisphere and below the plane in the southern hemisphere. Kahn and Woltjer suggest that this may be caused by the interaction of the corona of the Galaxy as it moves through the intergalactic medium. The densities within the galactic disk range from 3 atoms/cm^3 within spiral arms to a tenth of this value between the arms. It can be traced out to $R = 14$ kpc by which distance the density has fallen to 0·1 atoms/cm^{-3}.

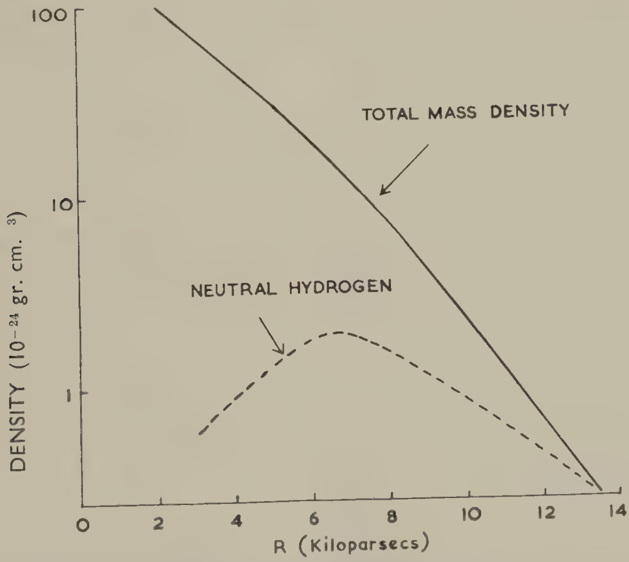


Fig. 5. The density of neutral hydrogen and the total density (stars plus neutral hydrogen) are plotted against the distance (R) from the galactic centre.

The optical rotation curve with additions from neutral hydrogen measurements can be used to construct a model of the distribution of mass density (stars plus gas) within the Galaxy using the following considerations. A star or gas cloud moving around the galactic centre is in a state of equilibrium where the gravitational attraction of the total amount of matter between it and the centre is balanced by the centrifugal force acting upon it. The mass of stars and gas within a given radius can thus be estimated from the observed rotation velocity for that radius. A recent model by Schmidt of the distribution of mass within the Galaxy is given in fig. 5. It shows the density of material in the galactic plane and can be compared with the averaged distribution of neutral hydrogen which was constructed from the observed longitude distribution of neutral hydrogen atoms. This distribution contrasts with that of the total mass density and suggests that neutral hydrogen and most likely the younger stars are more common in the outer parts of the Galaxy. It is perhaps worth noting here that the spiral structure observed in external galaxies is exhibited by only a small fraction of the mass of such galaxies. The bulk of the mass is in older less luminous stars which show no spiral structure and are concentrated toward the centre of the Galaxy.

5. THE CLOUD STRUCTURE OF SPIRAL ARMS

We have not yet discussed the way in which the hydrogen is spread throughout the spiral arms of our Galaxy. It had generally been considered as a result of optical studies that the interstellar medium consisted of gas clouds in motion relative to one another. Spectrum measurements of the hydrogen line confirmed this picture for they showed that the emission from any arm was spread in frequency due to the turbulent movement of the hydrogen in the line of sight. The observed random velocities vary from about 12 km/sec at $R=4$ kpc to 10 km/sec at $R=8$ kpc. Similar velocities are required by theory to support the different masses of material above the plane at these values of R.

The question of whether the neutral hydrogen is concentrated in discrete clouds with nothing in between or is irregularly spread throughout an arm, nowhere very dense and nowhere of zero density, was answered by observing the hydrogen-line absorption spectrum of the brightest radio sources. Here the emission from the source on its passage to the observer traversed a small number of clouds each absorbing radiation at a particular doppler-shifted frequency and giving rise to narrow absorption lines in the otherwise continuous spectrum of the source. In the case of the Cassiopeia-A radio source (see fig. 6) which lies in or beyond the Perseus arm, two heavily absorbing clouds in the Perseus arm and one in the Orion arm can be seen. However, if the neutral hydrogen was distributed in a random way throughout the arm the absorption spectrum would be similar to the emission spectrum because the absorbing cone beginning on the observer and ending on Cassiopeia-A would be a fair sample of the material in the arms. But the hydrogen appears as distinct clouds moving at quite distinct velocities within a range of ± 8 km/sec for a given arm.

An observation of considerable interest to astrophysicists which arose from the first hydrogen-line investigations was that the kinetic temperature of interstellar neutral hydrogen was rather more than 100°K—various observers have derived values between 120°K and 150°K. Theory, however, had shown that the temperature of the interstellar gas should lie between 30°K and 60°K. Although this discrepancy between observation and theory may appear small it demonstrates however that some large source of heat must have been neglected in the theoretical

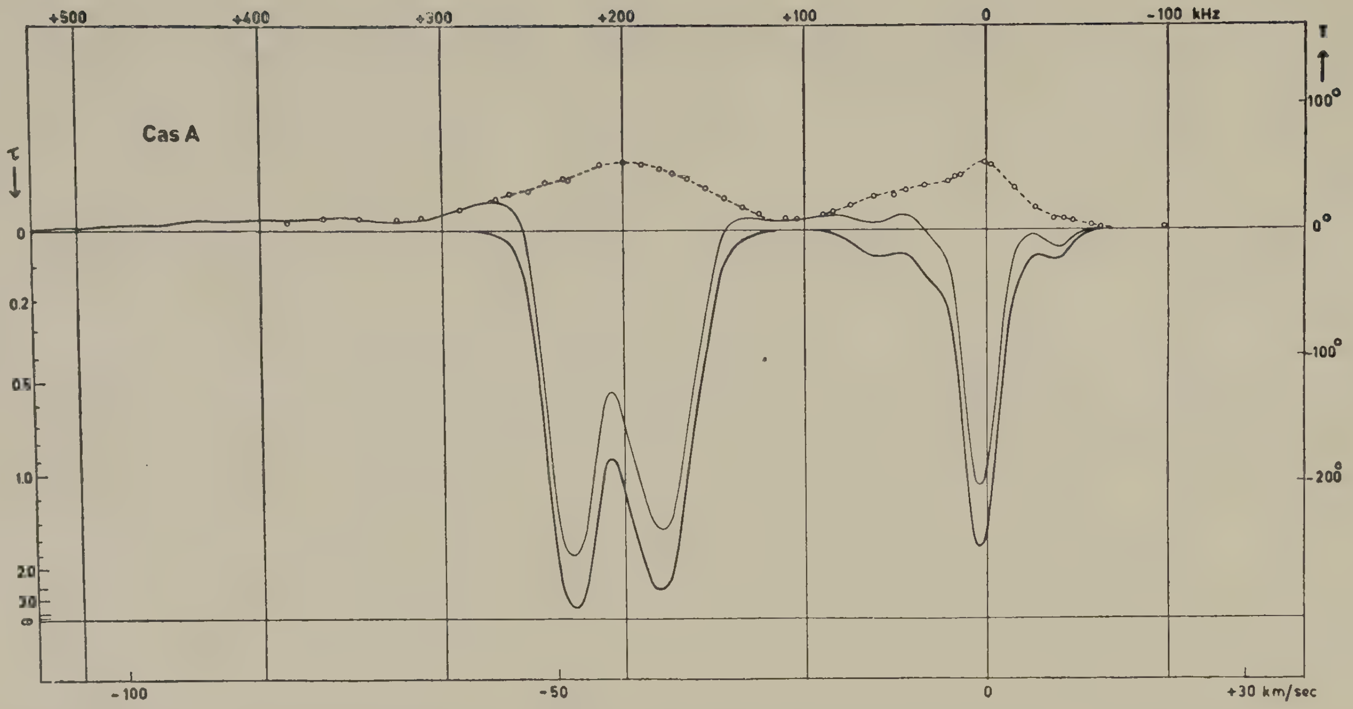


Fig. 6. The neutral hydrogen absorption spectrum of the Cassiopeia-A radio source is the heavy curve. The emission spectrum near the source is shown dotted. The single strong absorption line is caused by neutral hydrogen in the nearby Orion arm and the double line by neutral hydrogen in the Perseus arm. (By courtesy of C. A. Muller and the editors of the *Astrophysical Journal.*)

arguments. Kahn suggested that this heat originated in the collisions between clouds during their long journeys through interstellar space. During collisions the gas atoms are heated to temperatures in excess of 1000°K and afterwards cool to about 30°K due to radiation from metallic atoms and hydrogen molecules. After about 10^6 years a given cloud will suffer another collision and consequently interstellar clouds will have temperatures in the range a few thousand to a few tens of degrees Kelvin.

This still leaves us with the question of where did the clouds get their initial velocities from. Oort and Spitzer have made the very reasonable suggestion that the movements of the clouds are caused by their expulsion from the vicinity of hot O and B stars.

The low temperatures of some neutral hydrogen clouds expected on the theory of collision heating and subsequent cooling have been observed in practice. Values as low as 30°K were exhibited by clouds seen in projection against the hotter background clouds. Another method of determining an upper limit to the temperature of a neutral hydrogen cloud is to determine the width of its spectrum as seen, say, in the absorption spectrum of a source. The broadening is caused by the macroscopic turbulent movements within a cloud in addition to the thermal motions. Recent measurements at Jodrell Bank show the existence of some clouds with kinetic temperatures of less than 100°K which must be at an advanced state of cooling.

6. Star clusters and neutral hydrogen clouds

A glance at the sky on a clear night is sufficient to show that many stars are found in separated groups or clusters. These are sometimes classified into four different types. The *globular clusters* are compact assemblies of stars held together gravitationally. They are amongst the oldest members of the Galaxy having ages of 5×10^9 years, and are apparently devoid of any dust or gas which would have been swept out during their repeated passages through the galactic disk. Thus 21 cm emission is not likely to be detected from these objects which are spread like a halo around the Galaxy.

A second type of cluster to be found in this case within the galactic disk is called the *stellar association*. Many of these have been extensively studied by Ambartzumian and others. O-associations composed of O and B type stars have a low spatial density and cannot be gravitationally stable—indeed a number of such groups are found in a state of slow expansion. Ambartzumian also finds certain types of variable stars clustered together in dust clouds. These he calls *T-associations* after T-Tauri, the type star for this family of variables. Menon and van Woerden have independently studied the large expanding O-association in Orion which contains the Barnard Arc some 100 pc across. The neutral hydrogen density is $5/\text{cm}^3$ and the hydrogen mass is 3000 solar masses. 21 cm spectra suggest an expansion of about 12 km/sec which confirms the optical results. Although attempts have been made to detect neutral hydrogen from the I Lacertae association no hydrogen with a density greater than $1{\cdot}2\ \text{cm}^{-3}$ has been found.

A type of cluster which is found to be immersed in neutral hydrogen is called the *open cluster* to distinguish it in appearance from the compact and symmetrically shaped globular cluster. It is also gravitationally stable, but in contrast contains dust and young stars and lies within the disk of the Galaxy.

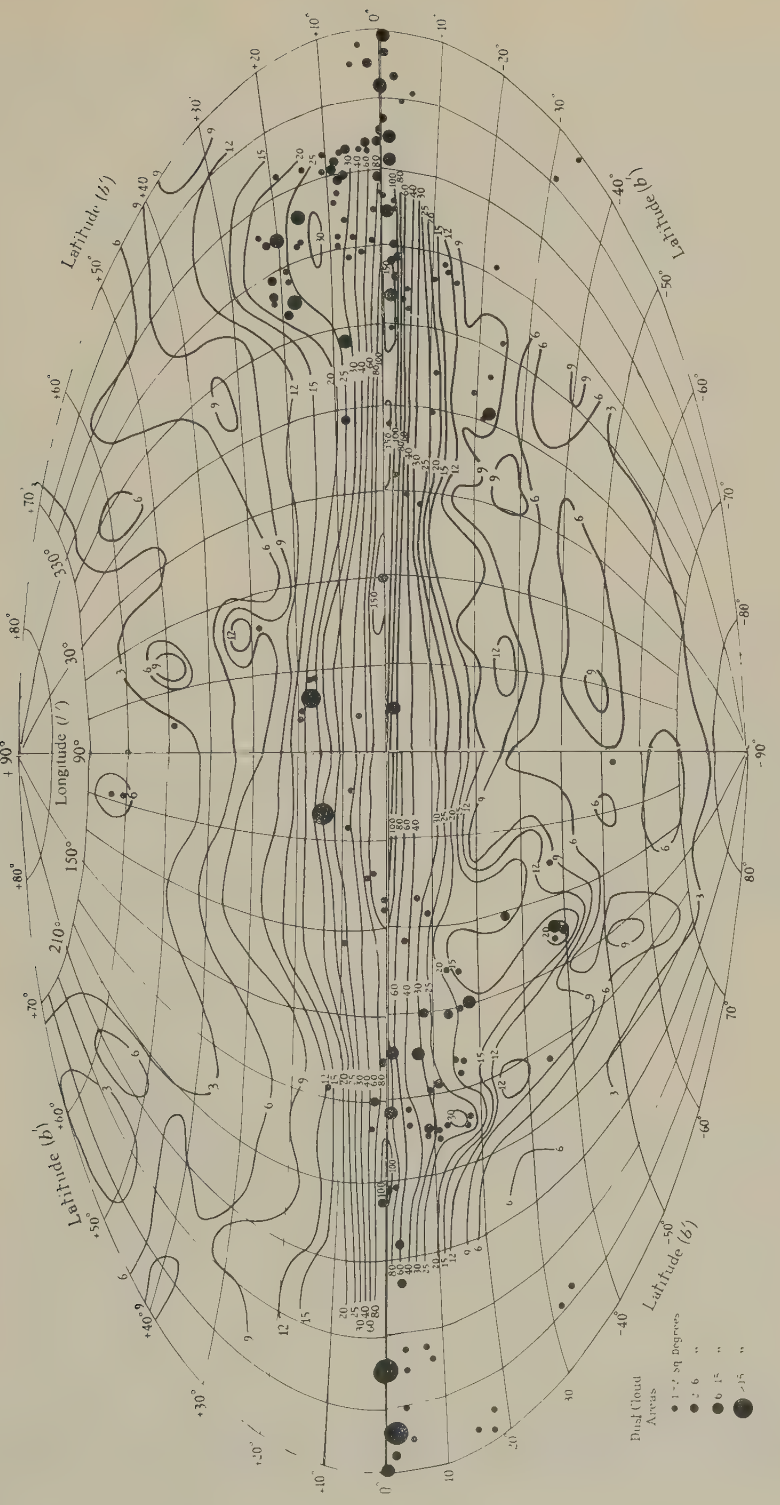


Fig. 7. The distribution of the total amount of neutral hydrogen in the line of sight plotted in galactic coordinates. The bulk of the hydrogen is concentrated in the galactic plane. The concentration of neutral hydrogen above the plane between $l' = 110°$ amd 190° and below the plane between $l' = 230°$ and 350° is the Gould Belt system. (By courtesy of the editors of *M.N.R.A.S.*)

The naked-eye Pleiades, Hyades and *h* and χ-Persei clusters belong to this group. The ages of these clusters can be estimated from the ages of the stars which constitute them. Drake has measured the amount of neutral hydrogen in the open clusters listed in table 1, and has discovered an interesting variation with age of the hydrogen mass and of the ratio of hydrogen mass to stellar mass. The younger clusters appear to contain a greater proportion of neutral hydrogen which he suggests is due to a progressive condensation of neutral hydrogen from the primaeval clouds into stars.

Table 1

Open cluster	Age (Millions of years)	Hydrogen mass (solar masses)	$\frac{\text{Mass of hydrogen}}{\text{Mass of stars}}$
h and χ Persei	3	7400	1·05
Pleiades	150	470	1·8
Coma Berenices	700	27	0·23
Praesepe	1000	26	0·08
M 67	5000	<18	$<0{\cdot}05$

A well-known star system not generally classified into one of the above groups is the Gould Belt of stars and gas. It was first recognized as a group of bright stars inclined at 20° to the galactic plane by Gould in the 1870's. When Shapley first estimated the dimensions of the Galaxy this bright system was shown to be a nearby one in which the Sun lay. Nearly all the B-stars brighter than 6 magnitudes were included along with the dust clouds and emission regions of large angular extent. Neutral hydrogen measurements have confirmed the existence of this region which can be seen in the contours of neutral hydrogen line-of-sight density in fig. 7. The Gould Belt system is superposed on the general background neutral hydrogen distribution which is concentrated along the galactic plane. Its greatest excursions from the galactic plane lie

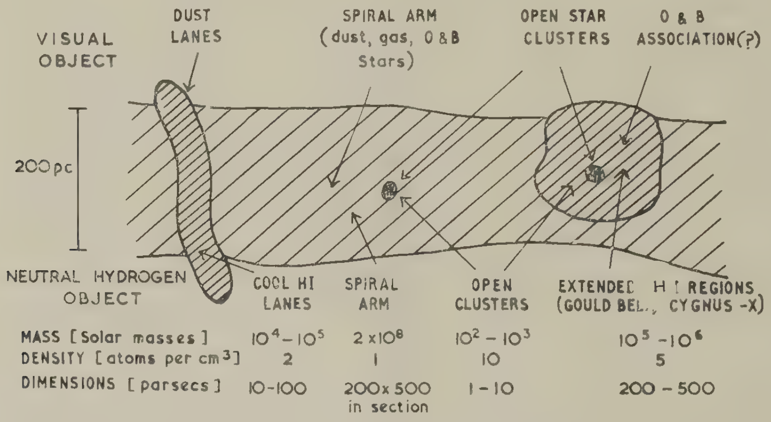

VISUAL OBJECT
DUST LANES
SPIRAL ARM (dust, gas, O & B Stars)
OPEN STAR CLUSTERS
O & B ASSOCIATION(?)
200pc
NEUTRAL HYDROGEN OBJECT
COOL HI LANES
SPIRAL ARM
OPEN CLUSTERS
EXTENDED H I REGIONS (GOULD BELT, CYGNUS -X)
MASS [Solar masses] 10^4-10^5 2×10^8 10^2-10^3 10^5-10^6
DENSITY [atoms per cm^3] 2 1 10 5
DIMENSIONS [parsecs] 10-100 200x500 in section 1-10 200-500

Fig. 8. A schematic diagram showing the masses, densities and dimensions of neutral hydrogen concentrations found in spiral arms. Some optical counterparts are also shown.

among the constellations of Taurus and Orion south of the plane ($l'=110°$ to $190°$) and of Scorpio and Scutum north of the plane ($l'=210°$ to $350°$).

Moreover the density of neutral hydrogen in the Gould Belt is unusually high relative to the star density and it may be supposed that this is a young system. This supposition is borne out by the observation that many of the stars are not more than 20 million years old compared with at least 200 million years for the stars in the local spiral arm. The Gould Belt system, with a mass of 10^5 solar masses and dimensions of 500 pc, is a giant system which is apparently gravitationally stable. It cannot be readily fitted into the classification outlined at the beginning of this section although it bears a resemblance to the rather less massive O-associations. It is probably similar to the large bright patches seen in the spiral arms of external galaxies.

Figure 8 summarizes the properties of the various stellar and neutral hydrogen regions seen in the spiral arms of our Galaxy. Their masses, densities and dimensions are included for comparison.

7. The galactic centre region

Radioastronomy has the particular advantage of being able to investigate the central regions of our Galaxy (which are inaccessible to optical astronomy because of the concentration of obscuring dust into the galactic plane). At the centre of the Galaxy lies a strong radio source clearly visible at centimetre and decimetre wavelengths which appears on the other hand as an absorption feature at metre wavelengths. These observations show that it must be an ionized hydrogen (thermal) region. Detailed investigations of this source by Drake at 3 cm wavelength revealed a bright region 10 min of arc in diameter surrounded by several fainter extended regions. Mills' survey at 3·7 m indicates that there is in addition a more extensive non-thermal region.

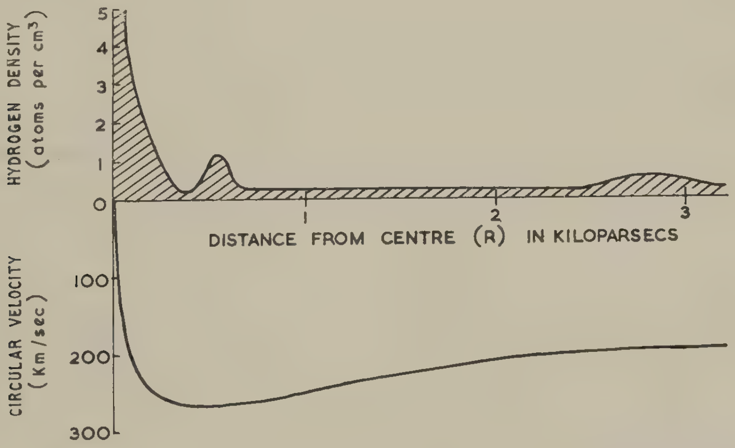


Fig. 9. The region near the galactic centre. The neutral hydrogen density along a radius from the galactic centre is shown as well as the circular velocity over the same interval. These plots are derived from the results of Rougour and Oort.

Neutral hydrogen measurements by Rougour and Oort lead to an even more involved picture of events near the centre of the Galaxy. Emission spectra show fast rotations within 20° galactic longitude of the centre and a clearly defined arm lies at $R=3$ kpc. This arm rotates at the normal galactic rotation velocity for $R=3$ kpc but at the same time is expanding outwards from the centre at 50 km/sec. Other minor arms are found nearer the galactic centre having outward velocities as great as 120 km/sec.

Within a distance of 600 pc of the centre is yet another rotating system showing rotational velocities of 265 km/sec at $R=550$ pc and 180 km/sec at $R=70$ pc (see fig. 9). Rougour and Oort explain this phenomenon in terms of the gravitational effect of a concentration of stars at the galactic centre. The density of hydrogen in their central regions rises from 1 atom/cm^3 at $R=600$ pc to 3 atoms/cm^3 at $R=100$ pc. The hydrogen density apparently rises to values of the order of 100 atoms/cm^3 at the centre itself where it is ionized by the concentration of stars.

A serious problem is raised by these observations of high expansion velocities in the central regions because it is readily computed that one solar mass passes out through the 3 kpc arm each year. Accordingly the central volume would be depleted of its hydrogen in 10^7 to 10^8 years which is a time short in comparison with the age of the galactic system (10^{10} years). The stars at the centre are unable to supply this amount of hydrogen so Rougour and Oort suggest that gas is funnelled into the galactic centre from the corona of the Galaxy. The understanding of this phenomenon is still in its infancy ; it is most likely associated with the unresolved problem of the production and maintenance of the spiral structure of the Galaxy.

8. Neutral hydrogen emission from external galaxies

The Large and Small Magellanic Clouds visible to the naked eye as large nebulous patches in the southern skies were the first external systems to be studied for their neutral hydrogen emission. These star clouds are our nearest extragalactic neighbours lying at a distance of about 60 kpc. They are known as irregular galaxies in contrast to the spiral galaxies like our own Milky Way system and to the elliptical galaxies which have a regular shape but no spiral structure.

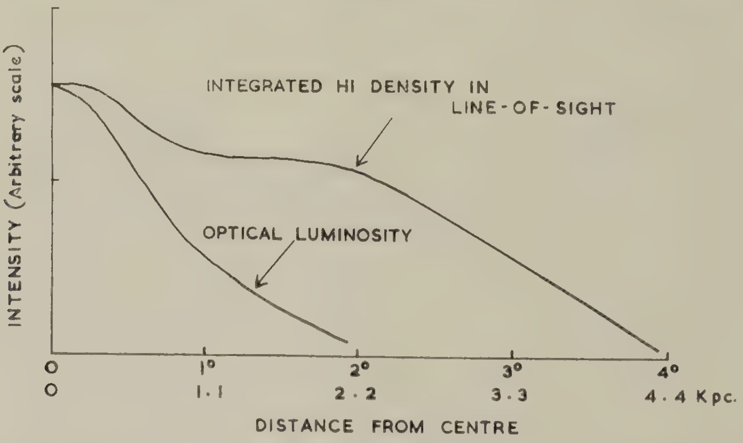


Fig. 10*a*. The neutral hydrogen line density and the optical brightness distributions along the major axis the Large Magellonic Cloud.

Kerr and his colleagues at Sydney plotted the distribution of neutral hydrogen throughout these Magellanic Clouds and found that it was more broadly distributed than the intensity of starlight throughout the systems (see fig. 10*a*). However one distinctive feature of the neutral hydrogen picture is the bridge between the two Clouds. The existence of this link may imply that the two systems are physically associated or it may be the result of disruptive tidal forces during a close encounter between them.

An important property obtained from hydrogen-line measurements for each cloud is its rotation curve which showed convincingly that both these systems were rotating. As in the case of our own Galaxy, these curves lead to an assessment of the total mass of each Cloud once the distance and the inclination of the system to the line of sight is known. The distance which has been accurately estimated allows the linear scale to be determined from the angular dimensions and also a knowledge of the inclinations allows the observed radial velocities to be converted into rotational velocities in the plane of the galaxy. For a regular system like a spiral nebula this angle can be readily obtained from the ratio of the major and minor axes on the assumption that the nebula is circular in plan. However, for irregular systems like the Magellanic Clouds this angle is subject to some uncertainty ; it has been estimated at 65° for both systems. The masses of the Small and Large Clouds are 2·0 and $5{\cdot}0 \times 10^9$ solar masses respectively.

A detailed study has been made of M 31, the giant spiral nebula in Andromeda, by van de Hulst and his colleagues using a 25 m diameter radio telescope. This galaxy is similar in structure and composition to our own Galaxy but is more massive by a factor of about 5. Both galaxies are denoted as of type S_b. The neutral hydrogen density within M 31 reaches a maximum at about 10 kpc from the centre and can be traced $2\frac{1}{2}°$ (30 kpc) in both directions along the major axis. This is beyond the distance to which stars can be found in any great numbers and consequently allows very useful new rotational velocities to be obtained. Again the mass distribution within the nebula was assessed from the rotation curve. Figure 10*b* gives the distribution of the neutral hydrogen density, the total mass density and the intensity of starlight for M 31. The neutral hydrogen mass is $0{\cdot}45 \times 10^{10}$ solar masses and the total mass is 48×10^{10} solar masses, each estimated for a distance of 630 kpc. The ratio of neutral hydrogen to total mass is 1 per cent, and is rather less than that for the Galaxy which has been given as 1·5 per cent. Both these estimates are subject to uncertainty and are probably not significantly different.

A further group of galaxies ranging in type from ellipticals, through spirals of types S_b and S_c to irregulars of the Magellanic type, have also been studied by the Dutch group. This survey was planned to show the way in which the neutral hydrogen density varied with galactic type. A rotation curve was obtained for the two large Sc nebulae, M 33 (the Triangulum nebula) and M 101 (the Pinwheel nebula). The calculated neutral hydrogen mass and total mass for each of the systems so far investigated are listed in table 2 along with their distances and galactic types. A systematic increase can be seen in the ratio of neutral hydrogen mass to total mass in passing from the ellipticals through the spirals to the irregulars. No neutral hydrogen mass greater than 0·4 per cent of the total mass was detectable in M 32, the elliptical companion to the Andromeda nebula ; while in S_b nebulae the ratio is 1 to 2 per cent, in S_c

nebulae 2 to 5 per cent and in irregulars 10 per cent or more. This confirms the suggestion often drawn from optical studies of the galaxies that irregulars are at an early stage of galactic evolution where they contain much hydrogen gas from which stars are condensing. At the other end of the evolutionary scale ellipticals show negligible signs of the interstellar gas from which stars long ago condensed.

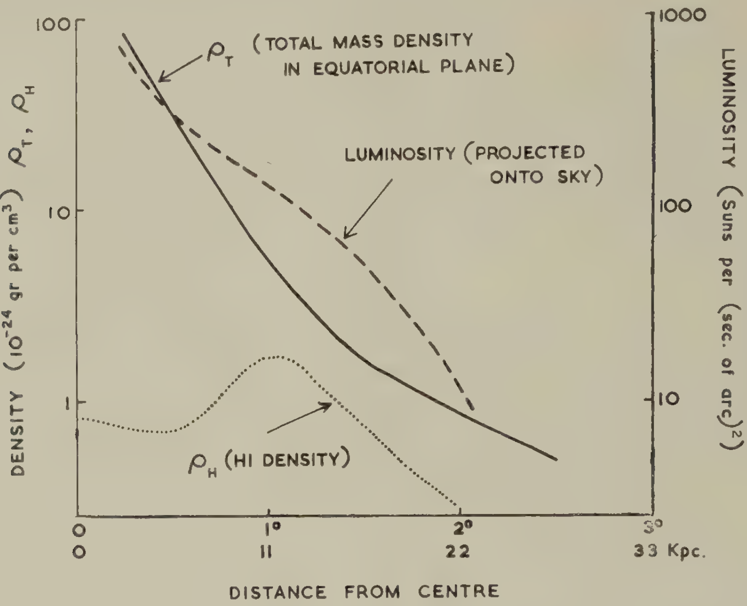


Fig. 10*b*. The density of neutral hydrogen and the total density (neutral hydrogen plus stars) as a function of the distance from the centre of the Andromeda nebula. Also shown is the brightness distribution across the nebula.

Table 2

Galaxy	Type	Distance (Kiloparsecs)	M_h = Hydrogen mass (unit = 10^9 solar masses)	M_t = Total mass (unit = 10^9 solar masses)	M_h/M_t (per cent)
M32	E_2	630	$\leqslant$0·06	14	$\leqslant$0·4
Milky Way	S_b	—	1·5	100	1·5
M31 (Andromeda Nebula)	S_b	630	4·5	480	0·9
M33 (Triangular Nebula)	S_c	630	1·0	18	5·5
M101 (Pinwheel Nebula)	S_c	2600	3·0	140	2·0
Large Magellanic Cloud	Irr	63	1·1	15	22
Small Magellanic Cloud	Irr	63	0·8	12	40
IC 1613	Irr	440	0·02	0·05	4

9. THE MAGNETIC FIELD OF THE GALAXY

There is strong observational evidence that magnetic fields exist within the Galaxy. However only one direct method of measuring the strength of these fields has so far been employed; this uses the Zeeman effect on the line radiation from neutral hydrogen. Under the influence of a longitudinal magnetic field a spectral line is split into left-hand and right-hand circularly polarized components which are separated in frequency in proportion to the strength of the field. The most suitable hydrogen-line features for this study are the narrow absorption lines of the strong radio sources and the line-of-sight magnetic field in the clouds which produce this absorption can be estimated directly as soon as the frequency split is measured.

A series of measurements at Jodrell Bank has failed to detect a longitudinal magnetic field down to a limit of about 5×10^{-6} oersted in the direction of the Cassiopeia-A source. Then using the conclusion from studies of the polarization of starlight that the magnetic field is parallel to the spiral arms, the general magnetic field of the Galaxy would appear to be less than or equal to 7×10^{-6} oersted. Such a small field, although at variance with some theories, is sufficient to line up the dust particles which produce the polarization of starlight, to hold the bulk of the cosmic rays in the Galaxy and to explain the observed isotropy of cosmic rays.

10. CONCLUSIONS

21 cm line studies give more information about physical conditions in space than measurements on any other frequency. They lead to estimates of density, temperature and velocity of the interstellar gas (and in consequence lead to a picture of the distribution of matter in space). In less than 10 years studies of neutral hydrogen in our own and other galaxies have greatly extended our knowledge of these objects, and have enabled us to probe regions inaccessible to optical astronomy. Even more new projects are in progress.

The large radio telescopes now in existence and under construction with beam-widths of 10 min of arc or less at 21 cm wavelength will soon be detecting galaxies which have red shifts of several percent of the velocity of light (200–300 million light years distant). Such a measurement can be used to check that the radio and optical red shifts are the same.

The magnetic field of the Galaxy can be directly measured by the Zeeman splitting of the hydrogen line. A low limit has already been set and more sensitive techniques are being used in an attempt to map the magnetic field throughout the Galaxy. These measurements have an important bearing on an understanding of the stability of the galactic spiral structure in which magnetic fields are believed to play a part.

SELECTED REFERENCES TO ORIGINAL PUBLICATIONS

The Mechanism of Emission
WILD, J. P., 1952, *Ap. J.*, **115**, 206.
PURCELL, E. M., and FIELD, G. B., 1956, *Ap. J.*, **124**, 542.
Techniques of Hydrogen-line Observation
MULLER, C. A., 1956, *Phillips Tech. Review*, **17**, 305, 351.
DAVIES, R. D., and JENNISON, R. C., 1960, Jodrell Bank Annals.

Spiral Structure of the Galaxy
van de Hulst, H. C., Muller, C. A., and Oort, J. H., 1954, *B.A.N.*, **12**, 117.
Oort, J. H., Kerr, F. J., and Westerhout, G., 1958, *M.N.R.A.S.*, **118**, 379.
Schmidt, M., 1957, *B.A.N.*, **13**, 247.
The Cloud Structure of Spiral Arms
van de Hulst, H. C., 1958, *Rev. Mod. Phys.*, **30**, 913.
Muller, C. A., *Paris Symposium on Radio Astronomy*, p. 360.
Davies, R. D., 1956, *M.N.R.A.S.*, **116**, 443.
Kahn, F. D., 1955, *Gas Dynamics of Cosmic Clouds*, p. 273.
Star Clusters and Neutral Hydrogen Clouds
Menon, T. K., *I.A.U. Symposium No.* 4, *Radio Astronomy*, p. 56.
Drake, F. D., *Paris Symposium on Radio Astronomy*, p. 366.
Davies, R. D., *M.N.R.A.S.*, **120**, in press.
The Galactic Centre Region
Rougour, G. W., and Oort, J. H., 1960, *Proc. Nat. Acad. Sci.*, **46**, 1.
Mills, B. Y., 1956, *Observatory*, **76**, 65.
Drake, F. D., 1959, *National Radio Astronomy Observatory*, Annual Report.
Neutral Hydrogen Emission from External Galaxies
Kerr, F. J., Hindman, J. V., and Robinson, B. J., 1954, *Aust. J. Phys.*, **7**, 297.
van de Hulst, H. C., Raimond, E., and van Woerden, H., 1957, *B.A.N.*, **14**, 1.
Volders, Louise and, van de Hulst, H. C., *Paris Symposium*, p. 423.
The Magnetic Field of the Galaxy
Davies, R. D., Slater, C. H., Shter, W. L. H. and Wild, P. A. T., 1960, *Nature*, **187**, 1088.

The progress of hydrogen-line research, and indeed radio-astronomy in general, may be followed in these conference reports.

(1) *I.A.U. Symposium No.* 4, *Radio Astronomy*, 1957, ed H. C. van de Hulst (Camb. Univer. Press).
(2) Proceedings of the Third Symposium on Cosmical Gas Dynamics, 1958, *Rev. Mod. Phys.*, **30**, 905–1108.
(3) *Paris Symposium on Radio Astronomy*, 1959, ed. R. N. Bracewell (Stanford Univ. Press).

The Nature and Properties of a Plasma

by A. VON ENGEL
Fellow of Keble College, Clarendon Laboratory, Oxford

1. INTRODUCTION

We have good reason to believe that much if not most of the matter in the universe is in the plasma state. This state can be described by comparing it with that of an ordinary gas, which is familiar to all of us, where large numbers of molecules move in all directions with different velocities along criss-cross paths, each sudden change of direction of movement being caused by a collision between two or more particles. But whereas in a gas all particles are electrically neutral, in a plasma charged fragments of formerly neutral molecules, positive and sometimes negative ions and electrons, take part in the random motion together with their neutral companions as well as with light quanta (von Engel 1955). In the gas of many stars, the sun, and certain apparatus in our terrestrial laboratories, charged particles predominate over neutral particles; these conditions approach a state called the fully ionized plasma (Spitzer 1956).

If we were to fill an imaginary box with a partially or fully ionized gas which has been produced by separating electrons from neutral atoms, and if all particles were uniformly distributed throughout the box, an observer outside this enclosure (assumed to have no walls) would find no electric field since such a ' neutral plasma ' contains equal numbers of positive and negative charges. If the observer, however, were to move sufficiently near to the boundary of the enclosure, say to distances of the order of the average distance between any two particles, or were to enter the box, he would observe very strong electric fields indeed which, because of the motion of the particles, vary in magnitude and direction. These fluctuating electric fields—the micro-fields—are a prominent feature of all gaseous matter in highly ionized state. They will be discussed later.

When a gas is being ionized the number of particles increases assuming a constant mass of gas. At large degrees of ionization the total pressure P can rise considerably above the original neutral gas pressure; this has been confirmed experimentally (Steenbeck 1939). Dalton's law thus apparently holds; with a mixture of gases

$$P = p_{\text{gas}} + p_{\text{ion}} + p_{\text{electron}}.$$

Before the last war the study of ionized and excited gases and of electric discharges in them was looked upon in this country as a fairly unattractive occupation, appropriate to times past and of little interest to the modern physicist of the thirties. The ' official view ' then held was that the two great masters J. J. Thomson and J. S. Townsend and their pupils had studied this field so exhaustively as to leave no major problems unsolved. In fact in those days the understanding of electric phenomena in gases in terms of fundamental processes had hardly started while nuclear physics was just becoming fashionable. The science of ionized gases has recently re-appeared on the academic stage under the name of ' plasma physics ' and is classed now as a ' U ' subject.

This change in outlook has not come quite unexpectedly. The developments in the study of the ionosphere, the advances in astrophysics, the discovery of thermo-nuclear effects (a field in which for the first time atomic and nuclear physics meet intimately), a large number of technical applications and the start of space research, have given a new stimulus to many frustrated theorists and experimentalists.

It is always difficult to trace with any certainty the originator of an idea. In England, W. Crookes (1879) was probably the first to describe the plasma state. He called it ' the fourth state of matter '. The term plasma, however, was first used by I. Langmuir (1929) in his investigations of a long uniform positive column of an electric discharge ' moulding ' so to speak ordinary neutral, excited and charged particles and light quanta which form the ionized gas. A mathematical definition of a plasma will be given in section 3.

A plasma is often the result of the action of an electric discharge in a gas. To produce it, energy is taken from a source which maintains a potential difference across the gas and the electric field drives a current through the gas. The steady state, where energy losses balance the energy input is fundamentally different from the thermal equilibrium which more easily lends itself to theoretical treatment; in practice it is difficult to realize because thermodynamic systems with zero losses can rarely be attained. In this ideal case the motion, interaction and transfer of particles from one energy state into another (excited atoms) is uniquely described by the temperature of the system.

According to a principle first enunciated by Klein and Rosseland, for every process going in one direction there must be a similar one going in the reverse direction, thus keeping the populations of particles in the various energy states the same, apart from fluctuations. For example, per unit time the number of atoms in the ground state which by absorbing light quanta are raised to a certain higher energy state is the same as the number of transitions from this energy state down to the ground state (accompanied by emission of quanta of that energy). Or, the number of atoms hit by electrons and producing positive ions and secondary electrons is the same as the number of ions recombining with free electrons—the recombination energy coming from the kinetic energy of the electron which happens to be near enough to neutralize the positive ion.

In the steady state the plasma populations of ions, electrons, excited particles and quanta similarly do not vary with time; but here energy is continuously fed into and lost by the system. For this reason the temperature is not, strictly speaking, a rigorous parameter fit for a description of such a plasma as a whole; though it is often convenient to apply it to the various populations. This is particularly so if the energy exchange between different groups of particles is restricted, as for example between gas molecules and electrons. Because of the large ion/electron mass ratio of 2000 to 500 000 elastic collisions can result in the formation of an ' electron gas '; the energy distribution of the electrons is thus often described by an *electron temperature*—equivalent to the average energy of a particle in terms of $\boldsymbol{k}T$. Similarly *ion temperatures* and *excitation temperatures* have been used to describe the corresponding distributions.

However, for many problems it is necessary to know the type of distribution which is not always Maxwellian. For example ' very hot ' transient plasmas (Francis 1960) contain often a considerable number of fast (' run away ') electrons as a result of the electric field acting on the electrons which travel through a tenuous gas whose collision cross-section decreases with increasing energy.

2. PRODUCTION OF A PLASMA

We shall distinguish between arrangements which give a steady plasma and those which give transient plasmas. An example of a steady plasma is shown in fig. 1. An arc discharge in an air stream or any other gas at atmospheric or at lower pressure is set up between two carbon or water-cooled metal electrodes.

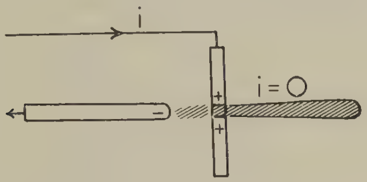


Fig. 1

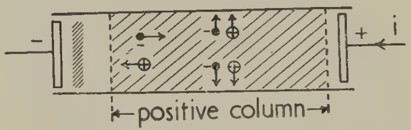


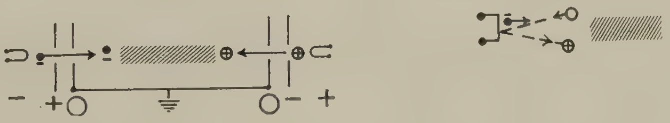
Fig. 2

Fig. 1. Quasi-neutral continuous or discontinuous plasma jet at high gas pressure. Moderate but equal gas and electron temperatures. Zero net current in plasma.

Fig. 2. Steady glow-discharge plasma, at low gas pressure. Gas at room temperature, electrons at high temperature. Finite current through the plasma.

Through the hole in the anode a hot plasma jet of considerable length (1 –10 cms) escapes which represents a neutral plasma. The net current flowing through the jet is zero and so are the macroscopic electric and magnetic fields in and around it. The arc currents are usually between 10 and 1000 amps; gas temperatures up to 20 000°K can be obtained with charge densities up to $10^{16}/\text{cm}^3$.

Figure 2 shows a steady uniform positive column of a glow or arc discharge in molecular or rare gases at either low or high pressure. It is distinct from the plasma in fig. 1 in that it carries a current and thus there are axial and radial components of the electric field and a tangential component of magnetic field. There are many variants of this arrangement; for example cooled constrictions are used in order to obtain high current-densities along a short length of the current-path, which results in higher electron densities and gas temperatures.



(*a*) (*b*)

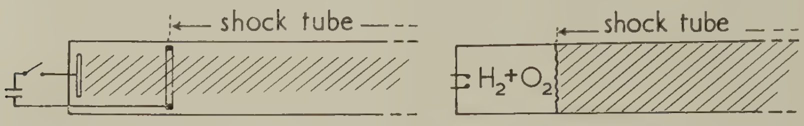
Fig. 3. Steady ' synthetic ' plasma. Gas is practically absent, and hence, if the interaction between the charges is strong, this is a fully ionized plasma.

(*a*) Electrons and ions produced separately, and the streams move in opposite directions.

(*b*) Electron-emitting surface acts as positive ion emitter when struck by neutral atoms (e.g. W and Cs), so both streams move in the same direction.

Attempts are being made at present to study steady fully ionized ' synthetic plasmas ' which are essentially free of neutral gas. Fig. 3*a* shows that this could be obtained by emitting electrons from one electrode and positive ions from another electrode, so that the corresponding beams of charges move in opposite directions. If the number-density of charges in the two beams can be made sufficiently large and the neutral density sufficiently low, collisions between the charges would be predominant and a fully ionized gas plasma should form. The same object can be achieved by emitting the two beams with opposite charges from the same end of a tube, fig. 3*b*; the ions can be formed, e.g. by surface ionization (von Engel 1955, Knechtli and Wada 1961).

A fairly simple way of producing transient travelling plasmas is by means of shock tubes. Figure 4*a* shows an apparatus which consists of a section in which a transient gas discharge of high intensity is produced by discharging a condenser through rare or molecular gases at low pressure. The adjoining shock tube which contains the same gas at the same initial pressure carries the shock wave with the associated plasma down the tube where its properties can be investigated. Another method (fig. 4*b*) of producing shock wave and plasma is by exploding electrically a mixture of hydrogen and oxygen in a combustion chamber separated by a thin breakable diaphragm from the shock tube, which is evacuated or contains a suitable gas. Both of these methods project a strongly luminous and very hot plasma along the tube.



(*a*) (*b*)

Fig. 4. Plasma pulses produced in shock tubes. Zero net current in plasma, gas pressure low or moderate or vacuum.
(*a*) Combustion produced by intense gas discharge; the initial gas pressures in discharge tube and shock tube are equal.
(*b*) Ionization, excitation and pressure wave produced in combustion chamber separated by a diaphragm from the shock tube section.

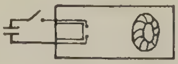


Fig. 5

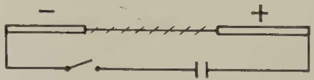


Fig. 6

Fig. 5. Plasmoids, which are vortex rings originating in the adsorbed gas released from the insulator carrying the spark gap. A current pulse is observed along circular axis of torus.

Fig. 6. Vapour plasma obtained by fast fusing of a thin metal wire.

A toroidal plasma (of vortex ring or ‘ smoke ring ’ shape) can be produced by using a tube containing what is really a special type of spark plug (fig. 5). Finally, a metal vapour plasma may be produced by passing large currents through thin fuse wires (fig. 6); melting occurs preferentially in certain sections of the wire (as pictures taken with high-speed cameras show), so that the plasma is a series of disconnected blobs whose investigation presents considerable difficulties.

3. General physical properties of the plasma

A plasma (fig. 7), although of course gaseous, has certain similarities with matter in non-gaseous forms. If the charges, instead of moving at random, were frozen in and assembled in a regular pattern, the similarity between such a ‘ low temperature plasma ’ and a crystal would be obvious. If, however, the charges consisted of positive and negative ions (instead of electrons) a comparison of this 'ionic plasma' with an electrolyte, or with ionic crystals, would be appropriate. Moreover, certain properties of semi-conductors can be compared with ionized gases containing 'impurity atoms' which give the plasma special properties.

How closely a plasma resembles a crystal can be expressed by the ratio C of the potential energy of a plasma-charge e separated d cm from its closest like neighbour (the crystal energy) and the kinetic (random) energy of the electrons. For a cube $d = N^{-1/3}$, where N is the number of charges per cm³,

$$C = (e^2/d)/\boldsymbol{k}T = e^2N^{1/3}/\boldsymbol{k}T \tag{1}$$

T is the temperature of the *electron gas* in °K. In general $C \ll 1$ and so the crystal-like behaviour of the plasma cannot be expected to be very pronounced (Cook and McEwan 1958). A comparison with electrolytes, however, seems to be more fruitful.

It can be shown that there exists in the plasma a characteristic distance, corresponding to the lattice parameter d, which describes the size of a 'plasma cell'. Let us assume that the cell size is large compared with the diameter of the ion so that we can suppose the plasma charges to be distributed uniformly. Consider a positive charge in such a plasma. It will attract electrons and thereby increase the electric field in its neighbourhood; it will slightly repel the heavy positive charges, which reduces the field at greater distances from it. Further re-arrangement of the other surrounding charges will lead to a still stronger field decrease, certainly more rapidly than the inverse square of the distance. Roughly speaking the field is confined to a finite region; outside this cell the field is essentially zero. This picture can be exploited quantitatively. In order to find the approximate range over which a field E exists we move all electrons from a region D of the uniformly fully ionized plasma (fig. 8) to the right of the vertical line a. On the left of this boundary there is now a sheath of positive charges of constant density N. In this case Poisson's equation reads for $x > 0$

$$\frac{\mathrm{d}^2V}{\mathrm{d}x^2} = +4\pi eN \tag{2}$$

and by integrating it twice we obtain (with the boundary conditions $E = 0$ at $x = D$ and $V = 0$ at $x = 0$):

$$E = -4\pi eN(D - x) \text{ and } V = +2\pi eN(2Dx - x^2) \qquad \text{(3) and (4)}$$

Thus the work W to move an electron through the space charge from $x = 0$ to $x = D$ is $W = 2\pi e^2ND^2$. If this is done by the thermal energy of the electron gas it can be equated to the kinetic energy of an electron in one direction, $\frac{1}{2}\boldsymbol{k}T$. Therefore the 'shielding distance D' (in cm) at which the field has reached zero value, is

$$D = (\boldsymbol{k}T/4\pi e^2N)^{1/2} \approx 7(T/N)^{1/2} \tag{5}$$

In ordinary discharges with $T = 10^4$ to 10^5 °K and N between 10^9 and 10^{14} cm^{-3}, D is between 10^{-1} and 10^{-4} cm. Thus the shielding distance is usually larger than the average distance ($N^{-1/3}$) between charges which is 10^{-3} to 10^{-5} cm. This is also true for partially ionized or non-isothermal plasmas; the differences are insignificant. If we had assumed spherical symmetry and a Boltzmann distribution of charge density ($N \propto \exp.(-eV/\boldsymbol{k}T)$), the calculation would have given a distance called a 'Debye-radius' whose value is about the same as D from (5).

Equation (5) shows that the larger N is, the stronger is the field necessary to separate the charges and the smaller the space D to achieve this. As T is raised the random velocity of electrons rises; this favours diffusion and separation and the distance D over which the now weaker field extends is increased.

When a plasma is bounded by an insulating wall or probe so that an equal number of charges of both signs are forced to arrive and combine to give zero current to the probe, the electrons keep the wall at a potential which is negative with respect to the plasma. Thus only the fast electrons will reach the wall together with an equal number of positive ions while the slow electrons are repelled. The wall field extends over a distance (‘ the sheath ’) which is of the order D.

It is apparent that the shielding plays a dominating role in plasmas. For this reason a plasma is often defined as an ionized gas in which $D \ll R$, R being the tube radius.

Fig. 7

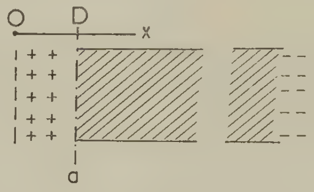


Fig. 8

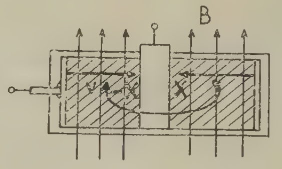


Fig 9

Fig. 7. Section of a plasma showing positive ions, electrons, excited and normal atoms and photons.
Fig. 8. Fully ionized plasma (shaded) and Debye's shielding distance D.
Fig. 9. Rotating plasma condenser.

Although the linearly averaged electric field $\boldsymbol{X}$ at a point in the plasma is zero, the averaged modulus $|X|$ is of the order $e/(d/2)^2 \approx 4eN^{2/3}$ (see 1), and in fact probably 3 to 5 times larger (Rompe and Steenbeck 1939). For $N = 10^{12}$ cm^{-3}, $|X| \approx 200$ V/cm, and for larger N the microfield produces an ‘ internal Stark effect ’ in the line emission spectrum. The atoms in the plasma ‘ see ’ a field X which varies with a frequency (noise) corresponding to the velocity of an electron crossing d; for our example it is 10^{12}/sec.

Some of the physical properties of a fully ionized plasma will now be derived from first principles. Several of the results can be obtained too from dimensional analysis (Brinkman 1959).

4. DIELECTRIC PROPERTIES

Since a steady uniform electric field will convey the charges in a plasma to the respective electrodes so that a direct current is set up in the circuit, there is no d.c. dielectric constant of plasma-condenser. The a.c. dielectric constant, the corresponding capacitance of a plasma and its dependence on the frequency of the field, however, can be understood.

When an electric field of amplitude X_0 and angular frequency ω acts upon a charge e of constant mass m in vacuo, the force is $eX_0 \sin \omega t = m(\mathrm{d}v/\mathrm{d}t)$, its velocity $-v = (eX_0/m\omega) \cos \omega t$ and the current density due to the oscillating motions of N charges per cm^3 is

$$\mathrm{j}_m = -eNv = -(e^2NX_0/m\omega) \cos \omega t \tag{6}$$

showing that $\mathbf{j}_m$ lags by $\pi/2$ behind X and is thus an inductive component. The other component $\mathbf{j}_c$ is due to the surface charges on the condenser plates and is of magnitude

$$\mathbf{j}_c = \frac{\epsilon_0}{4\pi}\,\omega X_0 \cos \omega t \tag{7}$$

Since $\mathbf{j} = \mathbf{j}_c + \mathbf{j}_m$, we find that with $\mathbf{j}/\mathbf{j}_c = \epsilon/\epsilon_0$ and $\epsilon_0 = 1$ (vacuum)

$$\epsilon = 1 - \frac{4\pi e^2 N}{m\omega^2} \tag{8}$$

i.e. the effective dielectric constant is reduced by charges oscillating in an evacuated condenser. This is to be expected because the inertial effects in mechanics are analogous to inductive effects in electric circuits. Thus, provided N is not too large, and changes of the local field X by space charges are small where

$$X = X_{\text{applied}} - f(N)$$

and furthermore, if either electromagnetic waves set up the field X, or electrodes which are so far away that charges cannot reach them, conditions can be such that the dielectric constant (8) can become zero or even negative. The former is the case when a vertically emitted electromagnetic wave reaches a height where the net electron concentration in the ionosphere is so great that it is reflected back to earth. This happens for frequencies of order one megacycle at a concentration of $\approx 10^5/\text{cm}^3$.

If, however, instead of a vacuum, we have a gas plasma, then the oscillating charges collide with gas molecules, are scattered, and acquire a random velocity. Hence their synchronous motion which originally lagged behind the field is now interrupted. After a collision a charge starts in general with an 'in phase' component of velocity. Thus work is done by the field on the charges. Their energy is transferred by collisions to the gas, whose temperature rises slightly. When charges are acted upon by an a.c. field X and their drift velocity v_d (the component parallel to X) is retarded by a force $= c_f \,.\, v_d$ where c_f is a constant 'coefficient of friction', then the balance of forces gives

$$m(\mathrm{d}v_d/\mathrm{d}t) + c_f \,.\, v_d = eX_0 \sin \omega t \tag{9}$$

Introducing the mobility μ, we have $c_f \mu X = eX$ and assuming Langevin's $\mu = (e/m)(1/\nu)$, ν being the collision frequency for a charge colliding with gas molecules, we find

$$c_f = (e/\mu) = \nu m \tag{10}$$

With this result the solution of (9) is

$$v_d = \frac{eX_0 \sin(\omega t - \phi)}{m\sqrt{(\nu^2 + \omega^2)}} \text{ with } \tan\phi = (\omega/\nu) \tag{11}$$

and the current density is

$$j = \{e^2 N X_0 \sin(\omega t - \phi)\}/m\sqrt{(\nu^2 + \omega^2)} \tag{12}$$

Now j lags behind X by a phase angle $\phi \leq \pi/2$. For $\omega \to 0$ the motion is simply controlled by the ion mobility. The equation for j contains a 'complex conductivity' $\sigma = e^2 N/m(\nu + i\omega)$; this implies a complex dielectric constant whose real part is

$$\epsilon = 1 - \frac{4\pi e^2 N}{m(\nu^2 + \omega^2)} \tag{13}$$

We conclude that collisions reduce the effective value of the dielectric constant. It has to be pointed out that (11) is an approximation which holds strictly only for ions in low fields; it can be used for electrons in those molecular gases where deviations from a constant mobility are not too large.

In plasmas of finite size the use of ϵ is often of doubtful value since it contains no real dipoles. Thus to apply ϵ, the amplitude of oscillation must be smaller than D or the size of the vessel.

An interesting application of the dielectric properties of the plasma is the 'hydro-magnetic capacitor'. This is essentially a small concentric cylindrical condenser filled with a gas at low pressure in which a dense plasma is established for a short time. By means of a magnetic field normal to the electric field the plasma is made to rotate so that it represents a gaseous flywheel of small inertia (fig. 9).

The dielectric constant of the linear system is easily calculated. Let the particle density ρ and velocity v be uniform throughout the gap. Then neglecting losses and equating the total field energy to the sum of electrostatic and kinetic energies (all per unit volume) we have

$$\epsilon\frac{X^2}{8\pi}=\epsilon_0\frac{X^2}{8\pi}+\tfrac{1}{2}\rho v^2 \tag{14}$$

Since the velocity v is produced by a magnetic field B normal to X

$$v=(X\,.\,\boldsymbol{c}/B) \tag{15}$$

where $\boldsymbol{c}$ is the velocity of light and X and B are expressed in e.s.u. and e.m.u. respectively. Note that the direction of v is independent of the sign of the charge and of the mass. Eliminating v, we obtain for the effective dielectric constant

$$\epsilon=1+4\pi\rho\boldsymbol{c}^2/B^2 \tag{16}$$

With argon at a pressure of 0·1 mm Hg, $\rho = 2{\cdot}3\,.\,10^{-7}$ g/cm³ and with $B=10^4$ gauss $\epsilon=2{\cdot}5\,.\,10^8$. With $X=3$ e.s.u. $=0{\cdot}9$ kv/cm the energy per unit volume from (14) is 10 joule/cm³. However, much higher energy densities have been obtained. A particular feature of this condenser is its low inductance and hence a very short discharge time ($\approx 10^{-5}$ sec), depending mainly on the length of the leads in the consuming system, can be expected. If the 'vacuum capacity' is 1 pF the dynamic capacity is $\approx 100\,\mu F$. When the plasma velocity is high ($\sim 10^6$ cm/sec) as well as its density, the centrifugal plasma pressure blows the magnetic field slightly outwards and ϵ rises by a factor of about 2.

5. Optical properties

Though the relation between dielectric and optical properties of gases has been known for a long time, their study has only recently been extended to plasmas (Alpher and White 1959). Thus from measurements of the refractivity and dispersion of a plasma, the degree of dissociation and ionization can be determined.

The refractive index of an ordinary gas in excess of that of vacuum, $(n-1)$, is equal to the specific refractivity times the gas density ρ, provided the density is not too high; for a gas mixture it is equal to the sum of the products of partial refractivities and densities. Again for a single gas the excess refractive index $n-1$ is proportional to the number concentration of particles N, and the mean polarizability α, which in turn depends on the oscillator strength for the transition in question, the frequency at which absorption occurs and the frequency of the light. This applies only to a gas whose particles have no permanent, but only an induced dipole moment. In a gas mixture, or in excited or ionized gases, the optical properties are assumed to be additive in character.

In general it is difficult to calculate the polarizability of a plasma, since simple expressions fail to describe α near absorption frequencies. In all cases where α is known n can be calculated from

$$n-1=2\pi\sum_{i}\alpha_i N_i \tag{17}$$

where N_i is the number of molecules/cm^3 of the i^{th} species. On the other hand the total refractive index can be determined interferometrically, and by using light of different wavelength and knowing their individual dispersions the fractions of the various components in the gas can be derived. If excited molecules are present (Dalgarno and Lewis 1957) the corresponding values α^* are to be used in (17) which are usually considerably larger than those of molecules in the ground state.

The refractivity of an ionized gas can be investigated similarly. Electrons make a major contribution to it. The result can be inferred from (8) with $\omega_p=4\pi e^2N/m$. This gives with zero damping

$$n-1\approx-\omega_p{}^2/2\omega^2 \tag{18}$$

and describes the dispersion where ω is the angular frequency of light and ω_p the angular plasma frequency. Since the latter is proportional $N^{1/2}$, the electron concentration and the degree of ionization can be derived from interferometric measurements. It is unnecessary to know the absolute value of n if measurements of n_1, n_2 at two different wavelengths λ_1, λ_2 are carried out. Thus N follows from

$$n_1-n_2\approx-A\,.\,(\lambda_1{}^2-\lambda_2{}^2)\,.\,N \tag{19}$$

where A is a known numerical constant.

Such measurements have been made on shock waves in argon at pressures of a few mm Hg, with interference filters and visible light beams normal to the shock. Electron concentrations of 10^{16} to 10^{17} electrons/cm^3 have been derived from interferograms. It is important to remember that this is the favourable range of application. Moreover, it is possible to derive in the same way the relaxation times, i.e. the time dependence of the degree of ionization.

The relations between optical and magnetic properties of plasmas (Vinti 1932) have hardly been applied yet, nor has use been made of the Faraday and Kerr effect.

6. ELECTRIC CONDUCTION

A plasma of infinite dimensions maintained at a given temperature and charge concentration possesses an electric conductivity σ which is determined by collisions between electrons and gas molecules, viz.

$$\sigma=\frac{j}{X}=\frac{eNv_d}{X}=eN\mu \tag{20}$$

provided the electrons carry the current and the contribution by positive ions is negligible. Here the symbol μ stands for mobility. In general the mobility of electrons μ^- is a fictitious quantity since it depends strictly on X and N. However, in many cases (20) gives the right magnitude and shows that σ increases with rising degree of ionization (e.g. increasing temperature) and, since $\mu^- p \approx$ constant, decreases with increasing gas pressure. A hydrogen positive column at $p = 1$ mm Hg with $N = 10^{11}$ electrons/cm^3 and $\mu^- = 4 \cdot 10^5$ cm^2/v. sec gives $\sigma \approx 6 \cdot 10^{-3}$ $(\Omega\ \text{cm})^{-1}$ or a resistivity of 160 Ω cm.

When the temperature of the gas is sufficiently raised so that thermal ionization occurs, the number of collisions between electrons and ions becomes prominent. For a fully ionized plasma an approximate value of σ can be easily derived from (20). According to Langevin's theory of mobility

$$\mu^- \approx \frac{e}{m} \frac{\lambda}{\bar{v}} = \frac{e}{m} \frac{1}{(3\boldsymbol{k}T/m)^{1/2}} \frac{1}{Nq} \tag{21}$$

where m and q are the mass of the electron and the collision cross-section for momentum transfer respectively and λ is the mean free path.

If an electron is strongly scattered by ions, i.e. through an angle $\pi/2$, when the ion potential energy $e^2/r_0 = 3\boldsymbol{k}T$ (where r_0 is the distance of closest approach), then

$$q = r_0^{\,2}\pi = \left(\frac{e^2}{3\boldsymbol{k}T}\right)^2 \pi \tag{22}$$

From (20), (21) and (22) we find

$$\sigma = \frac{(\boldsymbol{k}T)^{3/2}}{e^2 \,.\, m^{1/2}} \approx 4 \cdot 10^{-5} \,.\, T^{3/2} \text{ in } (\Omega\ \text{cm})^{-1} \tag{23}$$

A logarithmic factor of value $\frac{1}{3}-\frac{1}{6}$ which results from the shielding effect (section 3) is included in the numerical factor of (23). For a hydrogen plasma at $T = 10^4$ °K this gives $\sigma = 40 (\Omega\ \text{cm})^{-1}$. Note that σ is roughly independent of N because $\mu \propto 1/N$; N enters the shielding factor only logarithmically. A rise in plasma temperature *increases* σ; when T exceeds 10^6 °K σ becomes higher than that of metals at room temperature. Figure 10 shows how σ varies with N and T in a fully ionized hydrogen plasma. The conductivity for *electron circulation* is represented by σ^-.

This conductivity cannot be measured by inserting two electrodes in the plasma and passing a current through it unless the electrons which are removed from one electrode are replaced, e.g. by the other electrode. Otherwise positive and negative charges can only be extracted from the plasma at equal rates. Since in this case the (heavy) ions determine the rate of extraction a plasma conductivity σ^+ is found (fig. 10) which is at least ≈ 40 times smaller than that given by (23). The electronic conductivity $\sigma-$ has been measured by inducing currents in the plasma whereby electrons move on circular paths through their ion atmosphere (Lin *et al.* 1955), and the ionic conductivity σ^+ by a probe method (Sakuntala *et al.* 1960).

When a magnetic field acts on a completely ionized infinite plasma the conductivity depends on direction and is a tensor. If the current traverses the plasma normal to the direction of the field, the conductivity is found to be reduced by up to one-third of its zero field value (Spitzer 1956). The effect of the field is to change the energy distribution of electrons that are contributing to the current by reducing their average energy. This increases the number of slow electrons, the momentum transfer cross-section rises and the increase in effective collisions reduces σ^-, measured in the direction normal to that of the magnetic field. However, σ^- in the direction parallel to the field remains unchanged.

7. THERMAL CONDUCTIVITY

It is known that the thermal conductivity κ of a gas in a normal state is given by

$$\kappa = (\bar{v}\lambda/3) \, . \, \rho \, . \, c_v \tag{24}$$

where the first factor is the diffusion coefficient D, and ρ and c_v are the gas density and specific heat at constant volume respectively. When a gas is feebly ionized, the contribution to heat conduction by electrons is very small. Since $D \, . \, \rho$ is roughly proportional to $T^{1/2}$ and c_v increases with T, κ rises with T. However in molecular gases at $T > 10^3$ °K dissociation sets in and because of the energy required for it, $c_v = f(T)$ has a maximum beyond which the gas is atomic. Further peaks of κ occur when substantial amounts of energy are necessary for excitation (provided the process is confined to a narrow temperature range), ionization, etc. (fig. 11).

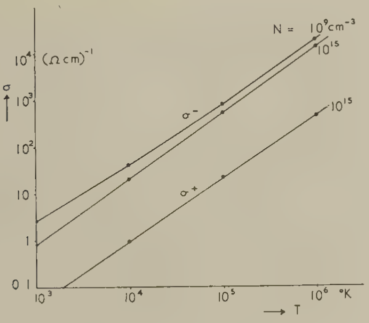


Fig. 10

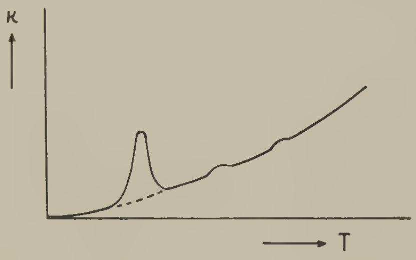


Fig. 11

Fig. 10. Electric conductivity σ of a fully ionized plasma as a function of temperature T and charge concentration N, in hydrogen. The curves labelled σ^- represent the electronic conductivity for different N, and that labelled σ^+ the ionic conductivity for H^+.

Fig. 11. General variation of the thermal conductivity κ of a molecular gas as a function of the temperature T. The first peak is due to dissociation, which happens at ≈ 3500°K for H_2 and ≈ 7500°K for N_2. The subsequent peaks are caused by excitation, ionization and other phenomena.

The heat conductivity of a completely ionized plasma is found to rise strongly with T. As expected κ is related to σ, the electric conductivity, by Wiedemann-Franz's law:

$$\kappa/\sigma = C_1 \, . \, T \tag{25}$$

and hence with (23)

$$\kappa = C_2 \, . \, T^{5/2} \tag{26}$$

The constant C_1 is for metals about $3k^2/e^2 = 2{\cdot}4 \, . \, 10^{-8} \mathrm{v}^2/°\mathrm{K}$, whereas for the plasma $C_2 = 4{\cdot}4 \, . \, 10^{-9} \mathrm{v}^2/°\mathrm{K}$. There are but few measurements (Burhorn 1958, Sekiguchi and Herndon 1958).

It might be of interest to compare the mechanism of unidirectional heat flow by conduction in a solid with that in an ideal plasma. Consider a metal rod between a hot infinite heat source at temperature T_1 and a cool infinite sink at T_2, both being electrically insulated. The heat flow from T_1 to T_2 proceeds by lattice vibrations (phonons) and electrons (de Broglie waves). Since the total electric current in the rod must be zero and electrons are forced to diffuse from T_1 to T_2, there will be an excess negative charge at T_2 and a corresponding electric field in the rod. Hence fast electrons of the distribution move down to T_2 against the field while slower electrons are accelerated by the field up to T_1; in the steady state the electric field is such that $i=0$. Since the exchange in energy between electrons and ions in the lattice is feeble, heat is transferred mainly by the electrons.

In a fully ionized gas the conditions are very similar; again electronic heat conduction predominates. The picture given resolves an apparent paradox: the total specific heat of a solid is to a first approximation independent of that of the ‘ electron gas ’ because in the Fermi-Dirac distribution of energy there is only an insignificant number of electrons free to contribute to the specific heat c_v. It would be erroneous to conclude from (24) that when $c_v \to 0$ κ is negligible since the electron density $N \propto \rho$ in metals is very high. As in the case of σ (section 6) κ is nearly independent of N because in (24) $\lambda\rho$ is approximately constant.

The thermal conductivity κ in a magnetic field B normal to the direction of heat flow is reduced by factor $1/(1+\omega^2\tau^2)$ where $\omega = eB/mc$ (Larmor-, gyro- or cyclotron frequency) and τ is the effective collision time.

8. Magnetic susceptibility

Consider a cylindrical metallic conductor which carries a current in an axial magnetic field B. Electrons are drifting slowly along the direction of B while moving simultaneously in circular paths around the field lines. The circular motion results from those electrons with a velocity component v_n normal to B, and occurs in a sense so as to reduce B. The circulating current would produce

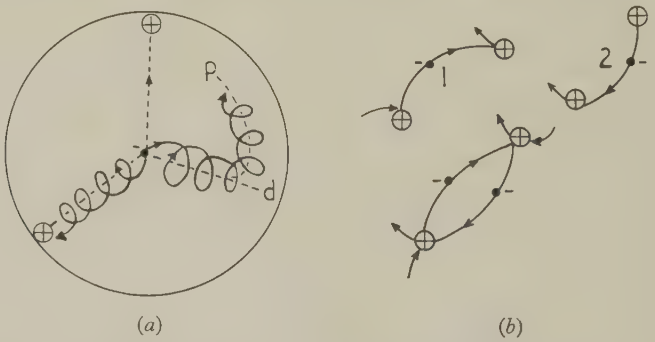


Fig. 12 (*a*) Circular motion of electrons in a positive column in a field B normal to the paper, producing diamagnetic (d) and paramagnetic (p) effects, or diamagnetic effects only when wall recombination with positive ions takes place.

(*b*) Compounding electron paths in a field B (normal to the paper) interrupted by collisions with atoms (circles with cross) which gives rise to diamagnetic moments. The two ‘ half-loops ’ 1 and 2 are equivalent to a closed loop as shown below.

in the volume a diamagnetic moment were it not for the reflection of electrons from the outer boundary of the conductor which gives rise to a circulation of charge in the opposite sense. The corresponding paramagnetic moment turns out to cancel exactly the diamagnetic moment (fig. 12*a*). This result is not unexpected: by applying a magnetic field the total energy of this system must remain the same if the electric field is to be zero.

If, however, a magnetic field acts on the plasma of a positive column at low pressure with both electrons and ions diffusing to the wall of the tube at the same rate, then those electrons which reach the wall are neutralized by ions and no paramagnetic component of circulation can develop; the plasma bounded by walls exhibits diamagnetic susceptibility (Rompe and Steenbeck 1939).

The dependence of the magnetic susceptibility χ of a plasma on gas pressure, magnetic field and the nature of the gas can be obtained either by calculating the sum of the areas given by the curved electron paths or the angular velocities about an axis. Assume that the radius of curvature ρ of the electron path in a field B is small compared with its mean free path λ (so that B/p is large). Obviously the motion will then occur as in vacuo. Since

$$\tfrac{1}{2}mv_n^2 = \boldsymbol{k}T_e \text{ and } mv_n^2/\rho = ev_nB \tag{27}$$

the magnetic moment M of N_e electrons/cm^3 is

$$-M = eN_e\rho^2\pi/t = \tfrac{1}{2}eN_e\rho v_n \tag{28}$$

where t is the time for one revolution ($t = 2\pi\rho/v_n$). From (27) and (28) the diamagnetic susceptibility/cm^3

$$\chi = -M/B = -N_e\boldsymbol{k}T_e/B^2 \tag{29}$$

If, on the other hand, ρ is larger than λ (B/p small), then the areas enclosed by circulating electrons are not full circles (fig. 12*b*) but two symmetric segments of total arc length 2λ, because randomness requires that for every electron moving on one side of the cylinder there must always be another symmetrically opposite moving in the opposite direction. The moment for N_e electrons/cm^3 is

$$M = -\mathrm{e}N_eAv_n/2\lambda \tag{30}$$

where v_n is the electron velocity and A the area of a double segment given by

$$A = \tfrac{1}{2}\rho^2(\lambda/\rho - \sin\lambda/\rho) \approx \lambda^3/12\rho \tag{31}$$

Substituting for ρ we obtain the magnetic susceptibility with e in e.m.u.

$$\chi = -\frac{e^2N_e}{24m}\lambda^2 \tag{32}$$

Figure 13 shows that the moment M rises as B is increased. This is because electrons drifting towards the wall follow a curved path (interrupted by collisions) which subtends a larger area. At even higher B, M falls as B rises because the path curvature is increased which reduces the subtended circular area. The variation of the susceptibility χ with B is shown in fig. 14. When B/p is small, $\chi \propto 1/p^2$ and independent of B. χ should be larger for the lighter gases, but in the case where there is a strong Ramsauer–Townsend effect ($\lambda = f(v)$) some averaging is necessary. For larger values of B/p, χ decreases as $1/B^2$ and is the larger the higher the electron temperature, i.e. the lower the gas pressure and the tube radius. Over the whole range of B, χ is proportional to the electron concentration and thus increases as the plasma current is raised (see also Tonks 1939).

Measurements of the susceptibility of positive columns (Steenbeck 1936) show that all the dependences are in agreement with theoretical predictions. However, the numerical constants differ by factors of 30 to 100. The reason for this disparity is the omission of the paramagnetic component. To allow for the radial electric field of the positive column which causes the paramagnetic susceptibility of the plasma, the ratio of the number of electrons neutralized on the wall to those reflected by the field of the wall charges must be known. It can easily be shown that only a small fraction of electrons is neutralized per cm. wall, viz. an amount $D_a/R^2 v_d$ where D_a is the ambipolar diffusion coefficient, R the tube radius and v_d the electron drift velocity, and that this fraction is of the order 10^{-2}. This brings the numerical values of the constants in line with theory.

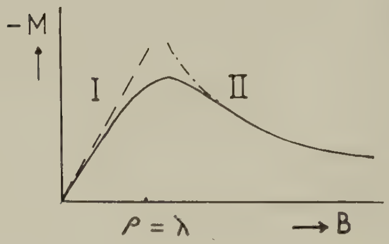


Fig. 13

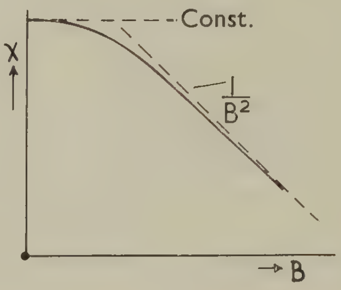


Fig. 14

Fig. 13. Diamagnetic moment M of a discharge plasma as a function of the field B showing a maximum near $\rho = \lambda$. The two branches I and II of the curve correspond to equations 30 and 29 respectively.

Fig. 14. Diamagnetic susceptibility χ as a function of the field B at constant pressure showing two limiting ranges.

In very strong magnetic fields other effects may take control like the change in the energy distribution of electrons or the variation of the radial distribution of charge density and the above results do not hold any longer.

9. DIFFUSION

Consider a feebly ionized plasma as it exists in the early stages of a positive column of a low pressure discharge at very small currents. Here the concentration of charges is so small that their electrostatic interaction is negligible. Therefore electrons, which are produced at a higher rate at the centre of the tube than near the wall, will diffuse through the neutral gas radially outwards. Their motion can be described by an electron diffusion coefficient

$$D^- = \frac{\bar{v}\lambda}{3} \qquad (33)$$

$\bar{v}$ and λ being the mean random velocity and mean free path of the electrons respectively.

As the concentration of charges of both signs is increased, e.g. by raising the current, electrostatic interaction between electrons and positive ions becomes stronger until both kinds of charges become so strongly coupled that they can only move with a common drift velocity. (Their random velocities are usually many orders of magnitude larger.) The drift velocity $v_d = -(D_a/N) \,.\, (dN/dr)$ can be described by an ambipolar diffusion coefficient

$$D_a = \frac{D^- \mu^+ + D^+ \mu^-}{\mu^+ + \mu^-} \tag{34}$$

whose value is between that of ions (D^+) and electrons (D^-). Since $\mu^- \gg \mu^+$ and $D^- \gg D^+$, (34) can be written approximately as

$$D_a = \frac{kT^-}{e} \mu^+ \tag{35}$$

where T^- is the electron temperature and μ^+ the ion mobility.

At still higher degrees of ionization the collisions between electrons or ions and molecules will become rare compared with the number of collisions between charged particles. The final transition to the self-diffusion coefficient of a plasma which is several orders of magnitude smaller than D_a, happens when the shielding distance is of the same order as the effective mean free path of ions in a fully ionized gas.

10. CONCLUSION AND ACKNOWLEDGMENTS

This introductory description of the physical properties of a plasma has attempted to show that with suitable simplifications and quite straightforward mathematical procedures we can describe the main features of the physical behaviour of a plasma, very largely in terms of classical theory. Of course it is by no means complete. For example, the containment problems, the oscillation and stability theories have already been discussed elsewhere (Chick 1960, Cole 1959 and Francis 1960) and have not been treated here.

I should like to acknowledge interesting and stimulating discussions with, and suggestions from, Drs. G. Francis and R. N. Franklin, and Mr. K. W. Arnold.

REFERENCES

ALLIS, W. P., and ROSE, D. J., 1954, *Phys. Rev.*, **93**, 84.
ALPHER, R. A., and WHITE, D. R., 1959, *Phys. of Fluids*, **2**, 162.
ANDERSON, O., *et al.*, 1959, *J. appl. Phys.*, **30**, 188.
ASCOLI-BARTOLI, U., and RASETTI, F., 1959, *Nuovo Cim.*, **13**, 1296.
BOSTICK, W. H., 1956, *Phys. Rev.*, **97**, 13.
BRINKMAN, H. C., 1959, *Physica*, **25**, 133.
BURHORN, F., 1959, *Z. Phys.*, **155**, 42.
CHICK, D. R., 1960, *Contemp. Phys.*, **1**, 169.
COLE, G. H. A., 1959, *Sci. Progr.*, **47**, 437.
COOK, M. A., and MCEWEN, W. S., 1958, *J. appl. Phys.*, **29**, 1612.
DALGARNO, A., and LEWIS, J. T., 1957, *Proc. roy. Soc.*, **240**, 285.
FRANCIS, G., 1956, *Hdb. Phys.*, **22**, 200 (Berlin: Springer-Verlag).
FRANCIS, G., 1960, *Ionization Phenomena in Gases* (London: Butterworths).

KNECHTLI, R. C., and WADA, J. Y., 1961, *Phys. Rev. Letters*, **6,** 215.
LANGMUIR, I., 1929, *Phys. Rev.*, **33,** 954.
LIN, S. C., RESLER, E. L., and KANTROWITZ, A., 1955, *J. appl. Phys.*, **26,** 95.
ROMPE, R., and STEENBECK, M., 1939, *Erg. ex. Naturwiss*, **18,** 257 (Berlin: Springer).
SAKUNTULA, M., VON ENGEL, A., and FOWLER, R. G., 1960, *Phys. Rev.*, **118,** 1459.
SEKIGUCHI, T., and HERNDON, R. C., 1958, *Phys. Rev.*, **112,** 1.
SPITZER, L., 1956, *Physics of Fully Ionised Gases* (New York: Intersc. Publ.).
STEENBECK, M., 1936, *Wiss. Ver. Siemens*, **15/2,** 1.
STEENBECK, M., 1939, *ibid.*, **18/3**, 94.
TONKS, L., 1939, *Phys. Rev.*, **56,** 360.
VINTI, J. P., 1932, *Phys. Rev.*, **41,** 813.
VON ENGEL, A., 1955, *Ionized Gases* (Oxford: Clarendon Press).
WILCOX, J. M., 1959, *Rev. mod. Phys.*, **31,** 1045.

The Author:

Has a doctor's degree in natural sciences of the Technische Hochschule Darmstadt and has spent many years at the Siemens Research Laboratories in Berlin. He has continued research in Oxford where he has taken a D.Sc. degree as a member of Christ Church. He was for several years lecturer in physics of Exeter College, an I.C.I. Fellow. He is in charge of a research group and teaches undergraduates at Keble College.

Accurate Voltage Reference Systems

by G. W. BROWN
Royal Aircraft Establishment

This article briefly surveys the various batteries and other electrical devices which are now available for use as accurate direct voltage references. Some information on their performance when subject to fairly severe mechanical vibration, is given, and a new temperature controlled voltage reference source using Zener diodes and germanium junction transistors is described.

1. Introduction

An accurate direct voltage reference is often required in aircraft and rocket instrumentation and this survey was carried out in order to determine which system showed promise of maintaining the greatest possible accuracy when subject to the accelerations, vibrations and temperature fluctuations present in vehicles of this type. It is well known that, in flight, these systems may be subject to peak accelerations of about 10 times that of gravity accompanied by fairly rapid changes in ambient temperature of several degrees centigrade. In addition, for one particular application, a fluctuating load current ranging from zero to a maximum of about 2 mA may be drawn from the reference source.

Most of the devices that have been considered up to date, are standard commercial products which were not originally designed for use under such adverse conditions, and the makers could give no guarantee of their performance in such an environment. As a result of this investigation however it is now believed that at least one system can achieve a voltage stability of three parts in 10^5 for periods of several hours under these adverse conditions.

The reference sources investigated may be broadly classified according to their basic principle of operation into three groups:

(*a*) Primary electrolytic cells.
(*b*) Gaseous glow discharge tubes.
(*c*) Solid state semiconductors.

There are a large variety of types within each group and during the short period of time available it has been impossible to investigate every possible product. Only the ones which appeared most promising to the author have therefore been studied and the salient characteristics of these components are listed in the table. A more detailed description now follows.

2. Primary electrolytic cells

The standard Weston cell is one of the oldest and in the right environment is still believed to be the best standard of direct voltage. Under carefully controlled conditions in a laboratory it can provide an absolute accuracy of about one part in 10^5 over a period of several years (Vinal 1950). Unfortunately its use is restricted by several factors. In the first place, it has a relatively high output impedance of about 1000 ohms, and a current of one hundredth of a microamp therefore produces a drop in output voltage of one part in 10^5.

This cell is also known to have a temperature coefficient of −0·004 per cent per °c at 20°c but this is only true if both electrodes in each limb of the cell change temperature by an equal amount. If the two limbs do not change temperature at the same rate, the cell exhibits a differential temperature coefficient about 10 times greater than this. When making accurate voltage measurements, therefore, great care must be taken to keep both electrodes at temperatures within about one hundredth of a degree centigrade of each other.

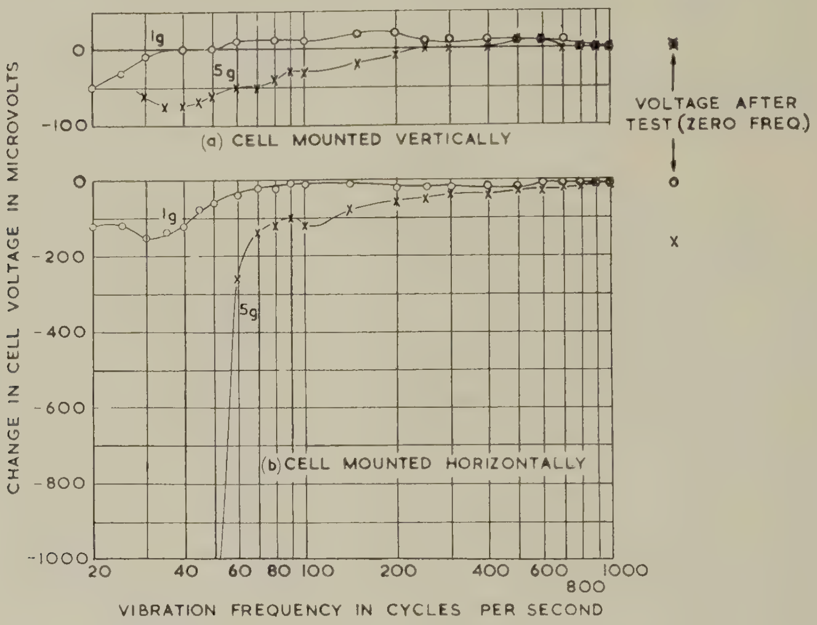


Fig. 1. Effect of vibration on voltage of saturated Weston cell.

In order to determine the effect of mechanical vibration, two of these cells have been subjected to vertical sinusoidal oscillations with peak values up to 5 times the value of gravity. Both cells behaved in a very similar manner and the graphs of fig. 1 show the change that took place in the direct output voltage of one of them as the vibration frequency was slowly changed with constant r.m.s. acceleration applied. Its relative voltage immediately after each frequency excursion, when the cell had come to rest, is also shown. The cell voltage was often affected for quite a long time afterwards, and sometimes never recovered to its original value.

These tests have shown that the cell voltage may change by about one part in 10^4 when mounted vertically and a vertical vibration of 5 g peak is applied to it. When mounted horizontally and vibrated vertically the effect is much worse as a peak vibration of 5 g produced a change of more than one part in 10^3 at the lower vibration frequencies.

Industrial type Weston cells with an unsaturated solution of cadmium sulphate have also been investigated. The voltage of these cells is not so predictable during manufacture as the saturated type and their voltage tends to fall more rapidly with use as the strength of the electrolyte changes.

Their output impedance and differential temperature coefficient is roughly the same as in the saturated type but their overall temperature coefficient is only 0·005 per cent/°c and is therefore nearly 10 times better than that of the saturated cell. In situations where it is inconvenient to correct for slow changes of temperature fluctuating over a range of a few degrees, these cells often have a considerable advantage.

Unfortunately when a few samples of these cells were vibrated their performance was worse than that of the saturated cell. It is believed that the Weston cell could be re-designed to improve its performance under vibration but its high output impedance would still be a serious disadvantage for this kind of application.

Another more recent type of primary cell often used as a voltage reference is the Ruben–Mallory cell described in detail by Vinal (1950). The RM3 size of cell has been chosen as a typical example. It is enclosed in a strong steel container and is much smaller and more robust than the Weston cell. It has quite a low output resistance of only a few ohms and a temperature coefficient of 0·002 per cent/°c, which is intermediate between the two types of Weston cell. Its differential temperature coefficient is unknown to the author but does not appear to be serious.

The performance of three of these cells when tested under vibration was very good with no sign of any resonances anywhere in the frequency band from 20 to 1000 c.p.s., even when vibrated at 10 g peak. Slow drifts of a few parts in 10^5 were sometimes observed but this may have been due to variations in the cell temperature which was not monitored during the test. It may also have been due to changing load conditions. It was not discovered until after this test, that the cell voltage displays a time lag of several minutes following a change in load conditions.

This time lag, and the fact that its impedance is not quite low enough, are the main objections to the use of this cell in the present application. Its performance is however quite promising when used under steady load conditions in the missile or aircraft environment.

3. Gaseous glow discharge tubes

A voltage reference tube of this type consists of two specially shaped electrodes mounted in a sealed glass envelope filled with various mixtures of inert gas at a reduced pressure of a few centimetres of mercury. If the direct voltage applied to its electrodes exceeds a critical value known as the ‘ striking potential ’ a discharge occurs and the gas ionizes, producing a characteristic glow at the cathode.

A resistance is normally connected in series with the tube to limit the discharge current but if limited too severely the ionized state cannot be maintained and the discharge current becomes intermittent. At the minimum value of current required to maintain a continuous discharge, the glow covers a comparatively small area of the cathode, but as the current is increased the glow spreads, with the current density and the voltage across the electrodes remaining almost constant until the whole of the cathode is covered by the glow. Up to this point the discharge is said to be ‘ normal ’ but with further increase of current the voltage starts to rise and the discharge is then said to be ‘ abnormal ’.

When operated in the ‘ normal ’ region the voltage across the electrodes is

Reference System		Absolute Voltage at 20°c	1000 Hour Stability per cent v	Temperature Coefficient per cent/°c	Operating Temperature range°c	Output Resistance ohms	Effect 10 g Peak Vibration per cent v	Volume cu. ins. Approx.	Weight ozs. Approx.	Remarks
Category	Type									
Wet Electrolytic Batteries	Saturated Weston Cell	1·01860 ±0·01%	±0·001	0·004 at 20°c	0 to +40	750 to 1500	0·1	3·6†	2†	†Unmounted cell
	Unsaturated Weston Cell	1·019 ±0·1%	±0·01	−0·0005 at 20°c	4 to +40	800 to 1200	0·1	1·2†	1·25†	
Dry Electrolytic Batteries	Ruben–Mallory Mercury Cell RM3	1·35 ±5±	−0·1	−0·002	0 to +70	1 to 10	<0·01	0·5	1·1	
	General Purpose Telephone Cell No. 6	1·6 ±5%	±2·5	+0·04	0 to +40	0·05		30	35	
Glow Discharge Tubes	Mullard M8098	85 ±2·5%	±0·1	0·005	−55 to +90	300	0·1‡	1·0	0·25	‡At maximum permissible vibration of 2·5 g. Ruggedized version of Mullard 85A2
	Mullard 83A1	83·8 ±1%	±0·1§	−0·003	−55 to +150	200		1·0	0·25	§Eventually settles down to a stability between 0·005% and 0·03%
Zener Diodes	Small British Zener Diodes	3·3 5·6 } ±5% 7·5	±0·003	−0·04 to −0·07 −0·02 to +0·02 +0·03 to +0·05	−40 to +100	10 to 350 1 to 90 1 to 20	<0·001	0·03	0·05	Output resistance depends on diode current. Two values given are for currents of 100 mA and 1 mA respectively
	American Zener Diode IN430A	8·4 ±5%	±0·003	±0·001	−65 to +150	11		0·35	0·8	
	Transistor Series Stabilizer 6·2 v Zener	6·5 to 9·0 Adjustable	±0·003	±0·003	+15 to +30	0·05	<0·0001	8	6	

largely independent of changes in the supply voltage and special reference tubes have been developed with very low temperature coefficients and very stable characteristics (Benson 1957). The Mullard 83A1 tube is one of the best reference tubes available at present and as will be seen in the table it is roughly comparable with the Weston cell in many of its performance figures. One defect is that its stabilized voltage drifts steadily with operating time, by about 0·1 per cent during the first 1000 hours of operation, but eventually settles down to a stability of less than 0·03 per cent after several thousand hours of operation.

No information on the effect of mechanical vibration on this tube is available at present but its performance in this respect is not expected to be very good. Special ruggedized versions of a similar tube, the Mullard 85A2 have been developed but none of these may be vibrated at more than 2·5 g without risk of failure. The Mullard M8098 is a typical example and has therefore been included in the table for comparison purposes. The output voltage of this latter tube is specified to vary by less than 15 mv r.m.s. when vibrated sinusoidally at 2 g peak acceleration.

4. SOLID STATE SEMICONDUCTOR DEVICES

Many British manufacturers are now making Zener diodes, a name commonly applied to all solid state semiconducting diodes that are intended to operate with a reverse voltage across them high enough to produce continuous breakdown. Their action is very similar to that of the gaseous discharge tube, and they are used in the same type of circuit with a series resistance to control the breakdown current.

These diodes usually consist of a small piece of very pure crystallized silicon, with a very thin layer of an impurity such as indium diffused into the crystal. This forms a thin depletion layer known as a *p-n* junction.

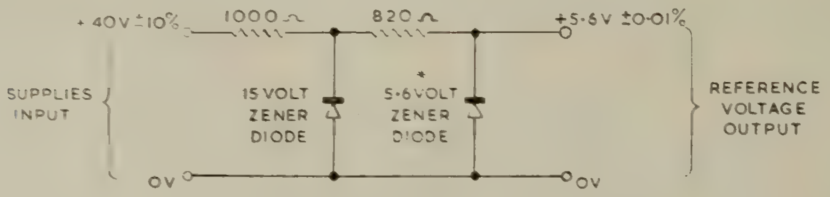
When a reverse voltage is applied to such a diode, almost the whole of the reverse voltage drop occurs across this layer. At low voltages, very little current flows through the layer but as the applied voltage is raised, the velocity of the free electrons in the layer increases until it becomes sufficient to produce hole-electron pairs when the electrons collide with atoms in the crystal lattice. Electron multiplication then takes place producing a rapid increase in diode current in an effect known as ‘ avalanche ’ breakdown (Dunlap 1957).

Breakdown occurs in this way when the width of the depletion layer is greater than the mean free path of the electrons. With narrower layers, a higher field intensity is required to produce breakdown, until with very narrow layers, the voltage gradient becomes sufficient to forcibly extract electrons from the atomic structure without the assistance of any collisions. This is the true “ field effect ” originally investigated by Zener, but his name is now generally coupled with all diodes which breakdown under either or both of these mechanisms.

Zener diodes are much smaller than gaseous discharge tubes and are readily available in a lower voltage range extending from 3·3 v up to about 100 v. Their dynamic resistance depends on many factors but is usually considerably lower than that of the gaseous discharge tube. Also the drift in their breakdown voltage with time is so small as to be difficult to measure and is definitely better than 1 part in 10^4 over periods of 1000 hours and is probably about 1 part in 10^5 in a good diode.

The stability of the reference voltage produced by a diode of this type is dependent on the voltage stability of the rectified mains or battery supplying the operating current. The improvement in stability achieved is usually expressed as a stability ratio, *S*, comparing the percentage change in stabilized output voltage with the percentage change in supply voltage. This ratio for the simple circuit consisting of one diode in series with a resistance is usually in the region of 0·03 to 0·01.

A considerable improvement in stability may be achieved by connecting two or more of these circuits in tandem as shown in fig. 2. The overall stability ratio then becomes approximately equal to that of the single stage raised to a power equal to the number of stages used. It is thus relatively easy to make the reference voltage independent of the supply voltage variations to any desired extent.



* Selected for low temperature coefficient

Fig. 2. Circuit diagram of Zener diode stabilizer.

Changes in temperature of the Zener diode usually produce an appreciable change in the stabilized voltage, but the effect can be minimized by a suitable choice of diode. An interesting feature, displayed by all makes of diode, is that their temperature coefficient is loosely related to their breakdown voltage, and varies from about $-0{\cdot}1$ per cent °c with a 3·3 v diode to $+0{\cdot}1$ per cent °c with a 15 v one, crossing the zero temperature coefficient point at about 5·6 v.

It is also dependent to a lesser degree on the actual temperature of the junction and on the current through it, becoming more positive as the current increases. Thus it has recently been demonstrated that in a batch of 5·6 v ±5 per cent diodes, 50 per cent had a temperature coefficient as low as 0·001 per cent °c somewhere within the range of 5 to 20 mA in an ambient temperature range of 20°c to 40°c.

If a higher breakdown voltage is required, two diodes with opposite polarity coefficients may be connected in series to give an equally low overall coefficient. A composite diode of this type is now available in America with a guaranteed temperature coefficient of less than 0·001 per cent °c over a temperature range of -55°c to $+150$°c.

When rigidly mounted, by potting the diodes and their associated components with araldite, in a temperature controlled metal block, their output voltage changed by less than one part in 10^5 when subject to 10 g peak vibration, in any plane and at any frequency in the range of 25 to 2000 c s.

For the particular application mentioned in the introduction, a lower output resistance of about 0·07 ohm is required and a slightly more complicated unit has therefore been developed (Robbins 1959, Chandler 1960). The circuit of this device is shown in fig. 3 and a photograph in fig. 4. It consists of a simple series transistor stabilizer using a 6·2 v Zener diode (OAZ203 in fig. 3) as the reference source.

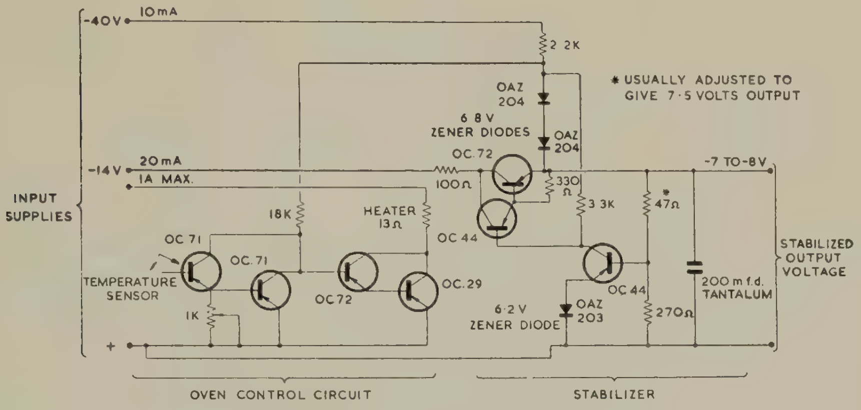


Fig. 3. Circuit diagram of transistor stabilizer with Zener diode reference.

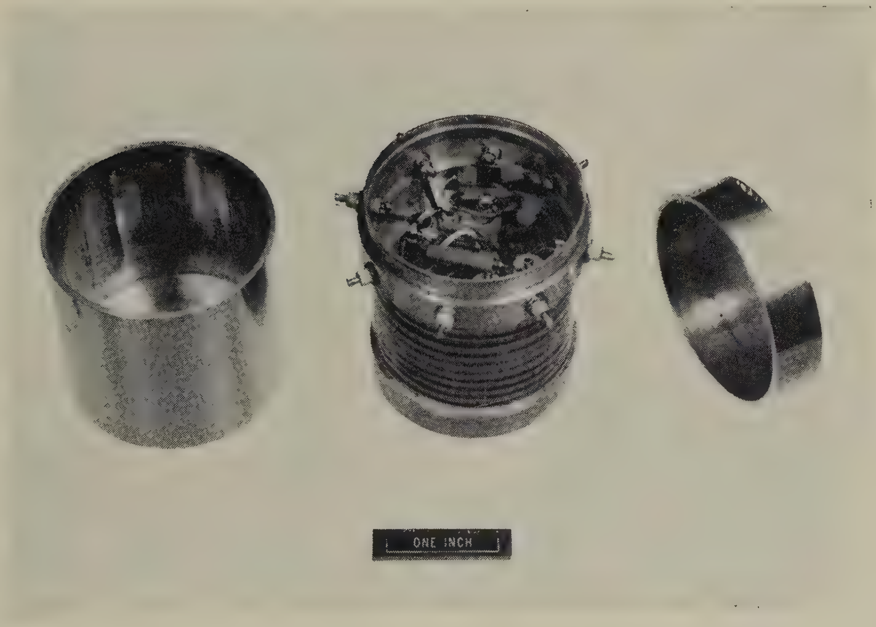


Fig. 4. Photograph of experimental model of transistor stabilizer.

The circuit of the stabilizer is shown in the right-hand half of fig. 3 and it operates in the following manner. Most of the 20 mA current from the −14 v supply flows through the OC72 transistor, the 47 ohm and the 270 ohm potential divider. The impedance of the OC72 is automatically adjusted to maintain an output voltage of 7·5 v across the potential divider and the tapping point on this divider then gives a voltage of approximately 6·2 v. The lower OC44 transistor compares this with the reference voltage of the OAZ203 Zener diode and any small change in the output voltage is amplified by the two OC44 transistors. The amplified voltage change is then applied to the control electrode of the OC72 and its impedance changes in such a way that the output voltage is pulled back almost to its original value.

The chief point of interest in this circuit is that the Zener diode is specially selected to cancel out, to a large extent, the negative 0·02 per cent/°c temperature coefficient of the transistor stabilizer. An overall coefficient of about 0·002 per cent/°c has been achieved in this way without much difficulty.

Rather than attempt to improve the overall temperature coefficient by more accurate matching, all the stabilizer components have been buried in a metal block which is temperature controlled to ± 0·2°c. If the ambient air temperature is always in the range 15°c to 30°c and any changes of temperature are fairly slow, i.e. there are no sudden draughts, it is possible to achieve this stability in block temperature with quite a simple transistor circuit, shown in the left half of the circuit diagram in fig. 3.

This circuit originally developed by Mullard Ltd. (Kemhadjian 1958) uses an OC71 germanium transistor as the temperature sensing element. The leakage current through this transistor, when operated with an open-circuited base electrode, is fairly high and increases rapidly as its temperature rises. This leakage current is amplified in a 3 stage d.c. coupled amplifier and controls, in reverse phase, the current through a heater winding wound round the outside of the metal block.

When the circuit is switched on, with the block cold, there is very little leakage current flowing in the sensing transistor and the OC29 transistor is switched hard on with maximum current flowing through the heater winding. As the block temperature rises the sensor leakage current increases and begins to switch off the current through the OC29. Due to thermal time lags the system usually overshoots and a slow damped oscillation occurs, but the block temperature settles down within a few minutes to a stable value in the region of 30 to 40°c depending on the value of the control resistance in the emitter circuit of the sensing transistor.

Several of these units have been built in the laboratory and the following performance has been achieved:

1. Output voltage =7·50 v.
2. Maximum output current =2 mA.
3. Output voltage stability:
    (*a*) With change of supply voltages. Minimum stability ratio with respect to the 40 v supply is 1500 and to that of the 14 v supply is 5000.
    (*b*) With change of load. One mA change in load current gives a maximum change of 1·0 part in 10^5 in output voltage.
    (*c*) With mechanical vibration. A 10 g peak sinusoidal vibration at any frequency from 25 to 2000 c.p.s., in any plane, gives a maximum change of 7 parts in 10^6 in the output voltage.
    (*d*) With a change of temperature. At a controlled block temperature of 35°c, a change from 20°c to 30°c in the temperature of the air surrounding the block gives a maximum change of 1·2 parts in 10^5 in the output voltage.

(*e*) With change of resistance in the output potential divider. The percentage change in output voltage is approximately one-sixth of the percentage change in either resistor.

Several versions of this type of stabilizer should be available in the commercial market in the near future.

ACKNOWLEDGMENTS

Much of the work described in this article was carried out at the Royal Aircraft Establishment in the Ministry of Aviation.

REFERENCES

BENSON, F. A., 1957, *Voltage Stabilized Supplies* (London: Macdonald).
CHANDLER, J. A., 1960, *Electronic Engng*, **32**, 78.
DUNLAP, W. C., 1957, *An Introduction to Semiconductors* (London: Chapman & Hall).
KEMHADJIAN, H., 1958, *Mullard Technical Communications*, **4**, 186.
KEMHADJIAN, H., 1960, *Reference Manual of Transistor Circuits* (Mullard Ltd.).
ROBBINS, A., 1959, Unpublished Ministry of Aviation Report.
VINAL, G. W., 1950, *Primary Batteries* (New York: John Wiley & Sons Inc.).

The Author.
Is a Principal Scientific Officer at the Royal Aircraft Establishment, Farnborough.

The Correspondence between a Current-carrying Coil and a Magnetic Dipole

by J. ALAN CHALMERS,
Reader in Physics, Durham Colleges
in the University of Durham

SUMMARY

The difficulties associated with the correspondence between the electric vectors **E** and **D** and the magnetic vectors **B** and **H**, when a current-carrying coil is considered as a magnetic dipole, are eliminated if it is considered, not as a simple dipole, but as a ' dipotential ' or dipole with a constant potential difference.

1. INTRODUCTION

In electrostatics it is possible to define two quantities, the electric intensity or field strength **E**, and the electric flux density or displacement **D**, related by $\mathbf{D} = \epsilon\mathbf{E}$, where ϵ is the permittivity of the medium. Similarly in magnetism and in electro-magnetism there are the two quantities, the magnetic field strength or intensity **H** and the magnetic induction or flux density **B**, related by $\mathbf{B} = \mu\mathbf{H}$, where μ is the permeability of the medium.

There is considerable disagreement among writers on the subject as to whether **H** or **B** should correspond to **E**, the other corresponding to **D**. Generally speaking, if magnetism is approached from the basis of permanent magnets, it is undoubtedly **H** which corresponds to **E**. But if the approach is by way of the forces on and due to currents, then there is justification for considering **B** to correspond to **E** and **H** to **D**.

The object of the present discussion is to show how this difference arises from the fact that a current-carrying circuit is not exactly equivalent to a simple magnetic dipole.

2. THE SIMPLE DIPOLE

A simple electric dipole consists of charges $\pm q$ separated by a distance d ; at any point near the dipole the vectors **E** and **D** have definite values, the ratio $\mathbf{D}/\mathbf{E}$ being ϵ, the permittivity of the medium.

If we consider media with different values of ϵ to fill, in turn, the whole of the volume round the dipole, then, since the charges on the dipole are taken to remain the same and since the charges represent the ends of lines of **D**, it follows that the value of **D** at any point is unaffected by the changes of medium. The value of **E** must therefore be proportional to $1/\epsilon$.

If we now consider the couple acting on a dipole placed in an electric field, with the length of the dipole perpendicular to the direction of $\mathbf{E}'$, the couple is $\mathbf{E}'qd$ where $\mathbf{E}'$ is the existing field strength. When the medium is altered, it is still the value of $\mathbf{E}'$ which determines the couple.

If we now consider the couple on one dipole in the field of another, with different media present, then $\mathbf{E}'$ is proportional to $1/\epsilon$ and so the couple varies as $1/\epsilon$.

For a dipole of magnetic material the same results are obtained, with **H** for **E** and **B** for **D.**

3. ELECTROMAGNETIC FORCES

There are various methods of approach to electromagnetic phenomena but they all lead to the same two fundamental results, which can be expressed in the form that a current-carrying coil behaves similarly to a magnetic dipole, both in producing magnetic fields and in the couple experienced in a magnetic field.

One fundamental law is that expressed by Ampère's theorem, that $\int H_s ds$ round a loop in space is equal to the current passing through the loop. Applying this to a current-carrying circuit in different media, it follows that **H** at any point is the same whichever medium is present. Comparing this with the electrostatic result, **H** must correspond to **D**.

The other fundamental law is that the couple on a coil is proportional to **B**; this cannot be verified directly since it is not possible to find a fluid medium whose permeability is sufficiently different from that of air, but the verification is obtained indirectly from the effects of electromagnetic induction. Again comparing with the electrostatic result, **B** must correspond to **E**.

The force between two current-carrying coils is proportional to μ, while the electrostatic case shows that this corresponds to $1/\epsilon$.

Thus the phenomena of electromagnetism appear to require **H** to correspond to **D**, **B** to **E** and $\mathbf{B}/\mathbf{H} = \mu$ to $\mathbf{E}/\mathbf{D} = 1/\epsilon$.

On the other hand, with true dipoles of magnetic material, **H** corresponds to **E**, **B** to **D** and μ to ϵ.

It is clear that, in fact, a current-carrying coil does not correspond exactly to a magnetic dipole.

4. PROPERTIES OF THE VECTORS

When one wishes to consider whether the electromagnetic phenomena or those concerned with true magnets give the correct correspondence with the electrostatic vectors, it is necessary to consider other properties of the vectors.

On crossing boundaries from one medium to another, **D** and **B** have normal components which are continuous, while for **E** and **H** it is the tangential components. Also in Maxwell's electromagnetic equations, it is the spacial derivatives of **E** and **H** which are related to the time derivatives of **B** and **D**.

These results agree with those obtained from true magnets rather than with those derived from electromagnetic forces and again suggest that a current-carrying coil is not exactly equivalent to a simple dipole.

5. THE 'DIPOTENTIAL'

The simple electric dipole is considered to consist of two charges separated by an insulator so that the charges remain the same, whatever may be the medium outside. It is, however, possible to think of another type of dipole in which there is a constant difference of potential maintained between the poles; we can imagine two small spheres connected by wires to the opposite poles of an electric cell. It is clear that the behaviour of these two types of dipole will differ when the surrounding medium is altered. We will call the dipole of constant potential difference a 'dipotential' and investigate its properties when the medium is changed.

The potential difference between two points can be obtained by finding $\int E_s ds$ along any line joining the points. If we take any line joining the poles of a dipotential, $\int E_s ds$ must be the same when the medium is altered and so $\mathbf{E}$ is unaltered at any point by a change of medium. Then $\mathbf{D}$ must be proportional to ϵ, and, since lines of $\mathbf{D}$ end on charges, the charges on the poles will be proportional to ϵ. This result could also be obtained by considering the dipotential as a condenser and using the well-known result that the capacitance of a condenser is proportional to ϵ; now, when the potential difference is kept constant, the charges concerned must be proportional to ϵ.

Considering now the couple on a dipotential placed with its length perpendicular to the direction of $\mathbf{E}'$, the couple is $\mathbf{E}'q'd$, where q' is the charge on the poles; but, as we have seen, q' is proportional to ϵ and so the couple is proportional to $\mathbf{E}'\epsilon$ and thus to $\mathbf{D}'$.

The couple exerted by one dipotential on another is therefore proportional to the value of $\mathbf{D}'$ due to the first. But this, as we have seen, is proportional to ϵ.

Comparing the dipotential with the dipole, we see that for the effects due to the dipole, $\mathbf{D}$ is unaffected by the medium, and for the dipotential it is $\mathbf{E}$; the couple on a dipole depends on $\mathbf{E}$, that on a dipotential on $\mathbf{D}$; and the couple exerted by one dipole on another is proportional to $1/\epsilon$, and for dipotentials to ϵ. In all, the roles of $\mathbf{D}$ and $\mathbf{E}$ are reversed in changing from a dipole to a dipotential.

6. COIL AS DIPOTENTIAL

Just as an electric dipotential as compared with an electric dipole shows a reversal of the roles of $\mathbf{D}$ and $\mathbf{E}$, so we would expect a magnetic dipotential as compared with a magnetic dipole to show a reversal of $\mathbf{B}$ and $\mathbf{H}$.

But comparing electromagnetic effects with those of a magnetic dipole we see just this same reversal of the roles of $\mathbf{B}$ and $\mathbf{H}$, and so we are led to the equivalence of a current-carrying coil to a magnetic dipotential, with the consequent correspondence in all cases of $\mathbf{H}$ with $\mathbf{E}$ and of $\mathbf{B}$ with $\mathbf{D}$.

This idea of a current-carrying coil acting as a source of magnetic potential is in complete agreement with ideas of the magnetic circuit with a 'magnetomotive' force provided by the current.

So long as we are dealing with one medium only, the results obtained for a current-carrying coil would be the same, whether it is treated as a dipole or a dipotential, and it is only when the medium is replaced by another that it is necessary to consider it as a dipotential.

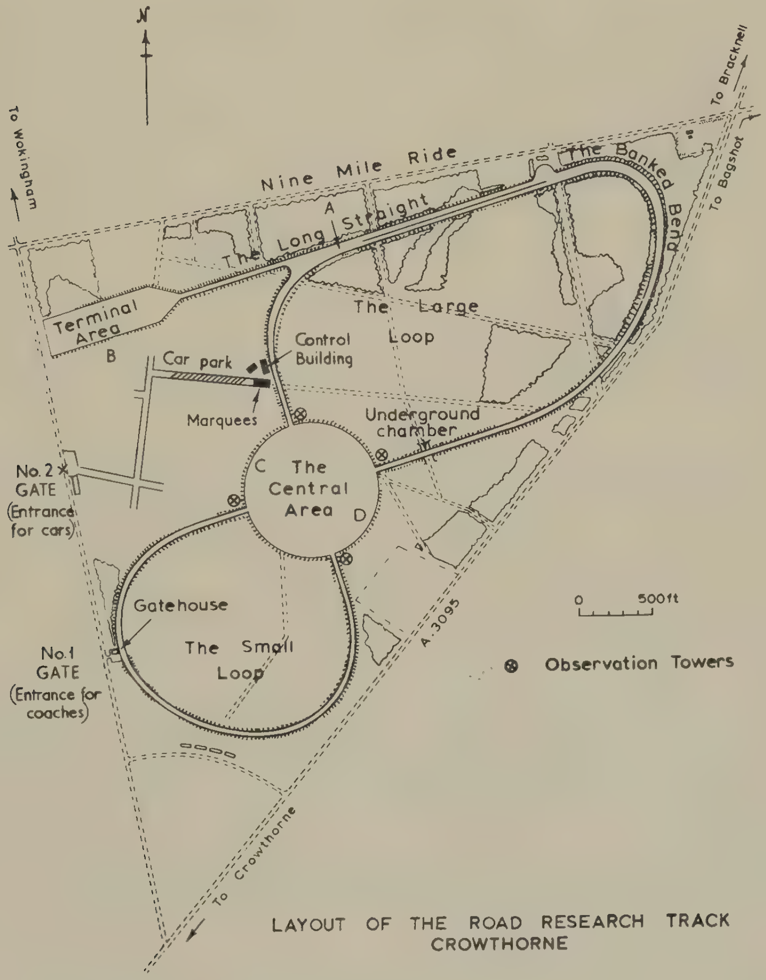
The Author:

Dr. J. Alan Chalmers, is Reader in Physics at the Durham Colleges in the University of Durham. He was born in 1904, educated at Highgate School and was a student at Cambridge and his Ph.D. was for radioactivity in the Cavendish Laboratory. His publications have been mainly in the field of atmospheric electricity, including two books on the subject.

DSIR Research Establishments

The Road Research Laboratory

In 1959, there were 261 216 road accidents involving an estimated cost of £219 million, 6520 persons killed with over 80 000 seriously injured and both of these numbers are increasing year by year, 8 606 049 vehicles licensed, and the number doubles itself in ten years, vehicles capable of ever greater speeds being put on the roads every day—all this shows the urgent need for practical road research facilities on a scale of such magnitude that they cannot be provided by ‘ private means and endeavours ’ but must be provided ‘ by public designation and expense ’. Did Bacon, when he wrote in the 17th century, foresee the establishment of DSIR ?



LAYOUT OF THE ROAD RESEARCH TRACK
CROWTHORNE

C.P.

At Crowthorne, in Berkshire, a new road research track has been built ; by 1964 the site will also house the entire Road Research Laboratory of DSIR—at present occupying two different centres at Harmondsworth, Middlesex, and Langley, Bucks.

The new track, which has cost about £500 000, is three miles long ; it is in the form of a figure eight with a large paved area (about 14 acres) in the centre, and contains a number of different test sections and various types of road surfaces.

The banked bend, designed for a safe speed of 60 m.p.h., enables investigations to be made in the long straight, on vehicles moving at 100 m.p.h.

In most measurements of skidding resistance it is essential that the road surface should be thoroughly wet and the long straight section has, therefore, been equipped with apparatus for spraying water over the test surface.

At one point there is an underground laboratory with a special glass panel let into the surface of the track. Photographic studies can thus be made of the contact areas of tyres as vehicles are driven over it, to examine the behaviour of tyres at the critical point of contact with the road surface and its effect on the skidding resistance of road surfaces. These studies, it is hoped, may point the way to a substantial decrease in the number of accidents due to skidding in wet weather.

The following note about the Engineering Services on the track was supplied by DSIR.

1. *Electricity supply, power and lighting*

The Southern Electricity Board brought in, by underground cable to a sub-station centrally placed on the site, the 11 000 v three-phase electricity supply. A 415 v supply from the sub-station provides the following facilities by means of underground ring cables :—

(*a*) *Service pillars :* 48 track-side service pillars, at 100 yard intervals along one side of the track only, provide power for experiments throughout the length of the track. Selected pillars have telephones and also socket outlets.

(*b*) *Track lights :* A simple lighting column beside each pillar enables the track to be used after dark. Each of these 48 mercury lights is also provided with a hand-operated switch to permit experiments which demand an unlit portion of track.

2. *Electronic vehicle control*

This first large-scale practical application of electronic vehicle control is based on installations in America by Radio Corporation of America and the General Motors Corporation. The system at Crowthorne has three main functions :—

(*a*) *Vehicle detection :* This is effected by having two-turn cable loops of vehicle size laid 9 in. under the road and placed with a 2 ft gap between each loop on the centre line of the ' Slow ' traffic lane. The loops are connected to terminal boxes at the edge of the hard shoulders. They will be fed with radio frequency low power supplies.

(*b*) *Vehicle guidance :* The installation comprises a guide wire system laid just below the road surface and on the centre line of the ' Slow ' traffic lane for the whole figure-of-eight circuit, and also 900 yards on the ' Fast ' lane terminating at the end of the skid track.

The audio-frequency electric supply will be connected to this circuit.

(*c*) *Collision prevention :* A second conductor is provided as a twin to the guide wire for use as antennae which will signal to drivers the relative speed and position of vehicles ahead.

The vehicle control system was installed to obtain practical experience in anticipation of electronic vehicle control facilities forming part of future public roads. In the course of his address to the visitors the Director of Road Research, Sir William Glanville, C.B., C.B.E., F.R.S., explained the main objects in designing the track.

Among the demonstrations were :—

Skidding resistance

(i) A car was equipped with an extra wheel so fixed that it was forced to slide sideways whilst the car was moving forwards normally. The skidding resistance was recorded on a moving paper chart. Lengths of up to two miles of road can be tested by this equipment.

(ii) For measurements of skidding resistance at high speeds the Laboratory has developed a small braking-force trailer which was towed by a car capable of making measurements at speeds of over 100 m.p.h. With the towing vehicle travelling at the desired test speed the trailer wheel was locked for 2 to 3 sec and a record of the force between the tyre and the road surface was obtained on a moving paper chart. This apparatus is used on motorways and other roads of that type and also on airfield runways where slipperiness can also be a serious problem in wet weather.

(iii) The skidding resistance of certain types of road surface changes markedly with wear. A portable skid resistance tester has been developed, consisting of a rigid pendulum the bottom end surface of which has been shod with a piece of tyre rubber. The pendulum swings in a vertical plane so that the rubber just grazes the road surface at the middle of the swing. The pendulum is withdrawn up to the horizontal position and then released ; after grazing, the amplitude of the motion is less than 90°—and this amplitude affords a comparative measure of the resistance.

(iv) Two cars came down the track at exactly the same speed and as near nose-to-tail as was consistent with safety. The leading car had ordinary tyres and the follower had tyres of high-hysteresis-loss rubber. At a given signal both drivers locked their wheels and both cars skidded to rest ; the stopping distance of the car with ordinary tyres appeared to be about 30% greater than that of the other. (See also B. E. Sabey, 1959, *Contemporary Phys.*, **1**, 56.)

Stability of vehicles in emergency braking

(i) In a car travelling at about 40 m.p.h. the brakes were applied locking all four wheels ; the car skidded to rest in a straight skid but turning the steering wheel produced no effect.

(ii) The front wheels only were locked and the result seemed to be as in (i).

(iii) In this case the back wheels only were locked, which sets up instability. The car started in a straight skid but rotation about a vertical axis set in under a couple which reached a maximum when the car was at right angles to the direction of skid.

Driver behaviour

One of the difficult problems of road research has been the study of driver behaviour, for example, when one vehicle follows another as in the exasperating queues of summer week-ends. In the system demonstrated, information from two moving vehicles, one following closely upon another, was transmitted by radio to a stationary receiving station. The distance travelled by each vehicle, the speed, the throttle position, and the amount of braking were recorded for both vehicles on the same moving chart. The investigation is aimed at determining the safe following distance ; it has been found that there is a time lag of 1¼ seconds, or more, between action by the leading driver and effective reaction by the follower—so what is the least distance of safe following in a queue moving at 6 m.p.h. ?

Traffic density detection

This was a demonstration of density detectors for queues of traffic. Two alternative methods of detecting the presence of vehicles—whether moving or stationary—which can be used to indicate where traffic is congested, were shown.

A simple computer which calculates the shortest route through a road network and in which particular roads can be closed to traffic was also demonstrated. Used in conjunction with density detectors and variable direction signs this device could route traffic away from roads which are temporarily blocked.

In addition to all this there were exhibits (*a*) showing experimental layouts in connection with the design of roundabouts, (*b*) illustrating the proposed work on the guidance of vehicles in fog and (*c*) of graphs of work in connection with 'dazzle'—a state which is said to exist when the light intensity exceeds 4000 candelas.

Here was the inauguration of a full-scale attack upon one of the most challenging problems of our time—and here also was practical physics ranging from coefficient of friction through photometry and radio to computers. Is it recruiting a sufficiently large number of young scientists of the right calibre to make the road death rate curve flatten out and go downwards ?

The Fire Research Laboratory

On 4th November, 1960, Lord Hailsham, the Minister for Science, opened new laboratories at the Fire Research Station, Boreham Wood. They include chemical engineering, hydraulic and radiation laboratories.

The Station is controlled by the Joint Fire Research Organization, set up in 1946 by the Department of Scientific and Industrial Research and the Fire Offices' Committee. The urgency of its work is brought out by the loss, each year, of about 700 lives, and material losses which last year reached a peak of £44 million ; a serious feature is the very large factory fire.

The work of the Station is organized in five sections—

1. *The Operational Research and Intelligence Section* deals with about 120 000 fires a year, attended by Local Authority Fire Brigades. Detailed studies are made of specific hazards—e.g. fires caused by oil-burning appliances, and by the domestic ironing board.

2. *The Ignition and Growth-of-Fire Section* studies the physical aspects of ignition, combustion and heat transfer, and the factors—particularly wind, ventilation, fuel and shape and size of compartment—which control the growth of fire and its spread within and between buildings.

3. *The Building Materials and Structures Section* studies building materials under fire conditions to obtain basic data for the design and protection of structures. Beams up to 24 ft in length can be subjected to prolonged temperatures ranging up to 1200°c.

   This Section also carries out, at the request of manufacturers, tests on structural elements and spread-of-flame tests on building materials.

4. *The Chemistry and Chemical Engineering Section* is concerned with industrial fires and explosions involving gases, vapours, dusts, and unstable materials not classified as explosives. Studies are also being made of the means of extinguishing fires in flammable liquids, the use of vapour-phase inhibiting agents, and the possibility of extinguishing fires in buildings by filling them with inert gas.

5. *The Extinguishing Materials and Equipment Section* concentrates on the properties of, and methods of applying, extinguishing agents such as water, foam, and dry powder. It also studies the automatic detection of fires and the performance of detection systems.

V.T.S.

Contemporary Reading

The Report of the Research Council of the Department of Scientific and Industrial Research (HMSO, Cmnd. 1365, May 1961, 4*s.*) mentions the steps being taken to coordinate large scale research. Three low-power research reactors are planned; for Scotland, Lancashire, and the London area. Cosmic-ray research will centre on Leeds, hydrogen bubble-chamber work (associated with the 25 000 Mev CERN proton-synchroton) on Imperial College, and optical spectroscopy on Manchester. One-year research studentships (so far only in chemistry) will give the ' II, 2 ' graduate a chance of probationary research; and unfulfilled promise may curtail research careers under grant. Efforts to attract back promising scientists who have migrated westwards, and to encourage a flow of research fellows in this direction from abroad, are being made. DSIR Reports have the same catholic spirit as the ' Boys' Own Paper ' of the Gordon Stables days, with its motto *Quicquid agunt pueri nostri farrago libelli;* this one records the successful fruition at the Building Research Station of a project for a wheelbarrow.

In the June issue (p. 381) Dr. F. A. Vick's course for schoolmasters at Harwell was mentioned, as an enterprise with incalculable possibilities for good. The idea was tried as long ago as 1957 in the U.S.A., when the Argonne National Laboratory gave a course on its activities to a group of high-school teachers. Published now by D. Van Nostrand and Co. as *An Introduction to Nuclear Science* (28*s.* 6*d.*) by Alvin Glasser, who represents 56 contributing authors, the course is interesting as an example of what teachers can be expected to take in. The lectures on basic nuclear physics, the metallurgy of uranium, radiation damage, biological effects, separation of isotopes, and radiochemistry, were reinforced by practical work. Not many of the experiments listed in the book are directly related to school teaching, but all would be valuable background for the teachers.

The need for re-equipping and maintaining the efficiency of science teachers is mentioned among all the statistics of *The Complete Scientist. An Enquiry into the Problem of Achieving Breadth in the Education of Scientists and Technologists* (O.U.P., 1961, 18*s.*). This is the report of a committee of the British Association under Sir Patrick Linstead's chairmanship, financed by the Leverhulme Trust Fund. One reason for this need is the age-distribution of the present teaching force; there must be many of us who graduated before the discovery of the neutron. Another is the maintenance of morale. A third good reason is that some other countries are doing it in a big way; this is well known, but it is often assumed that they have so far to go to catch up with us that they are merely working off arrears. Even if this were true, after some predictable time-interval t (or more probably δt) we shall be overtaken.

I really do do some contemporary reading of the normal kind, and much of it horrifies me in this context. *The Complete Scientist* was reviewed on the editorial page of " The Daily Telegraph " by T. F. Lindsay. Science, he said, is becoming a nasty word. Scientists, no doubt, are a good thing, but it is possible to have too much of them. And so on. This isn't Belloc's sly academic dig, or the Barlow Report's wild surmise. It is an influential journal in the year 1961, a notable supporter of the Administration, exerting its influence for the public weal by saying what it thinks the rather superior kind of public likes to hear. My guess would be that on this matter it is just about as much in touch with the smouldering misgivings of the public as are the Minister of Education and the Minister for Science.

The Complete Scientist analyses replies to questionnaires on syllabuses and programmes and performances, sent to schools, universities, industry, and the professions. It foresees the running down of the system unless more good teachers, of arts as well as science, enter the schools. This situation, it says, is not likely to improve while salaries in the teaching profession compare so unfavourably with those in other walks of life.

In a review (April, 1960, p. 329) I once doubted whether this was a major factor. The evidence here is that it is. But there is something deeper, a frustration within the job itself over amateurish lack of foresight at high levels, and over a national policy that prefers the cheapest action, whatever its ultimate cost; this is the underlying cause. To voice it openly is bloody-mindedness, bellyaching, and unscrupulously playing on your neighbour's conscience. To focus on one aspect, pay, is to have a case that all citizens understand and respect; and on which the authorities will give you a good, clean fight. Mr. Punch's picture, though, illustrates the basic and really crippling grievance.

The problems are not confined to this country, of course. One of the resolutions passed unanimously at the International Conference in Paris last year (the proceedings of which are now available as *International Education in Physics*, editors Sanborn C. Brown and Norman Clarke, publishers in Britain John Wiley and Sons, 36*s*.) was:

TOMORROW'S JAM

[According to UNESCO the USSR spends £37 a year on education per head of population, the USA £33 and Britain £9]

[*Reproduced by permission of Punch*]

" Efficient instruction in physics requires specialized teachers who can keep abreast of developments in a rapidly growing subject. We are alarmed at the present shortage of such teachers, particularly in view of the growing demand for physics education. The shortage is likely to become more acute in the years ahead.

In our opinion, steps should be taken to improve both the efficiency and the attractiveness of physics teaching as a profession. Insofar as the realization of these aims requires action by governments and universities, we recommend that these bodies should consider the following general conclusions:

1. In schools of secondary and higher levels, physics should be taught by physicists, that is, by men and women who have received a professional training in physics. Teachers must be encouraged to keep their professional experience up-to-date. The experimental nature of physics places an added burden on the teacher, and this must be recognized and adequately compensated by a reduction of his teaching hours and in other suitable ways.
2. To make teaching careers more attractive, improvements of salary and status are necessary in some cases, but most important are better conditions of work. For example, technical assistance and liberal provision of apparatus are vital, and facilities should be provided for all students at all levels to carry out experiments. Secondary-school teachers should have conditions in which they can feel that they form an integral part of the development of physical knowledge.

Universities and comparable institutions should accept their responsibility to establish close relations with secondary-school teachers, to assist in solving the problems of instruction in schools and to provide refresher courses. These courses would require extended periods of study leave for teachers."

This, it is true, is specifically concerned with physics; but though the field is narrow, the conclusions are world-wide in their relevance to it.

Lastly, the *National Physical Laboratory Report* for 1960 (HMSO, 1961, 9*s*. 6*d*.) shows that this institution itself seems to be developing an educational programme for its computers. There is a new Autonomics Department—autonomics being the science of designing self-organizing systems, which are intrinsically capable of improving their own performance. And there is a biological side of this studying the problems of human perception and reaction, in order to train scientists to educate *machines*. There must be a moral here relevant to the main theme of this article.

All the above was written and in proof before Mr. S. Lloyd's statement of July 25th; or some expressions would have been worded differently. I can remember the really serious crisis of 1931, when teachers were asked to accept greater salary reductions than others; but that was different, and if you had a secure job at all in those days you were better off than most. It seems to have set an unfortunate precedent for the present situation. This Journal can only deplore the Treasury's recent action, as one which will make the profession less attractive to prospective entrants and which flouts the expressed demand of the public as a whole for better conditions for teachers. It looks a strong-arm act; but £5½ m. in itself is merely a gesture, the little finger of Rehoboam extended with the thumb to the nose, a weak and flaccid provocation of dedicated servants. We cannot have an effective teaching force with a chip on its shoulder; the only sound long-term investment is to get rid of the chip and recruit more shoulders. When will it really sink in that the quality of the teachers and the teaching are matters for capital investment, and that even in bad times, or the worst hangover from a spending spree, you don't jeopardize your children's future by breaking into your capital? G. R. N.

ESSAY REVIEW

The Nucleus as it usually is.

The Principles of Nuclear Magnetism. By A. ABRAGAM. (Oxford University Press, 1961.) [Pp. xvi + 599.] 84*s*.

A considerable part of contemporary physics is concerned with the destruction, creation or rather uncivil treatment of nuclei as isolated individuals—selected individuals, as it were hand-picked to serve as projectile or target, or in some other highly improbable or unusual situation. This book deals with nuclei as they occur in bulk matter, that is to say in their *usual* situation. Very often such nuclei are extremely isolated having quite weak interactions with the rest of nature provided they do not try to creep out of their electronic shells. This is particularly true in respect of the orientation of the nucleus. A hydrogen nucleus in a sample of ordinary water for instance, in a magnetic resonance experiment, on the average continues to spin for about three seconds and may make something like one hundred million revolutions during that time without interruption. Stronger interactions may disturb this nuclear freewheeling but rarely (except in a few substances) more often than once a microsecond which on this sort of time-scale is an eternity.

As often happens in physics small interactions betray weightier matters. The weak interactions of the nucleus which are only observable with refined instrumentation can go some way towards telling us why water is fluid and ice is not, why copper conducts and copper sulphate does not, why some molecules are round and some are square, and so on. The Preface summarises the fields of application as follows, " structure of molecules, reaction rates and chemical equilibria, chemical bonding, crystal structures, internal motions in solids and liquids, electronic densities in metals, alloys and semiconductors, internal fields in ferromagnetic and antiferromagnetic substances, density of states in superconductors, properties of quantum liquids . . . ".

The spin angular momentum and the associated magnetic moment when present are here of prime importance. In classical spectroscopy the details of optical line spectra from sources placed in a magnetic field revealed the prescence of nuclear magnetism through very small effects. The *hyperfine* structure of optical spectral lines is as a flea on a much more powerful beast, the lines themselves which arise from electronic transitions. Recently the nuclear hyperfine effects have come back into their own in optical spectroscopy with Kastler's double resonance experiment, itself a form of nuclear resonance. However, we are concerned with the pure Zeeman effect for spins where all other interactions can be neglected in first approximation. For such comparatively free spins in a magnetic field we observe a resonant effect. An interaction with electromagnetic radiation only occurs within a range $\Delta\omega$ at a frequency ω where $\Delta\omega \ll \omega$ and hence we have a spectral line and nuclear magnetic resonance spectroscopy.

In dealing with the behaviour of magnetic spins it is interesting and instructive to note a reversal of the usual situation in quantum mechanics. Usually, in dealing with electrons for instance, the spacial part of the wave function is the most important. One sometimes feels that the spin wave functions are somewhat disembodied entities (where is spin space?) and are only put in to fiddle the selection rules and to pick out from a mass of possible wave functions the one that gives the right answer. This feeling is encouraged by the fact that one is able to leave out the spin functions entirely in a great deal of chemical physics with impunity. The study of spins has really come into its own in magnetic resonance. To one brought up primarily on ' space ' it is quite fascinating to see how the familiar procedures are applied to the manipulation of spins. This detailed and precise treatment of the quantum mechanics of spins for a great variety of interactions is beautifully illustrated in this book.

In the description of the interaction of spins with matter, that is with incoherent electromagnetic fields, we are on a much less sound ground. The spin dynamics of a liquid or a solid is a problem in quantum statistical mechanics and is very far from being satisfactorily solved. Necessarily the assumptions commonly used in the microscopic theory of irreversible processes must be carried over to the special problems of spin relaxation. The reader must not be surprised to find integrals extended to infinity, odd terms put equal to zero, differential equations solved by gross approximation and in some cases (perhaps?) the answer put in at the beginning, because this is usually done and because it gives more or less the right answer.

The systems studied here are almost ideally suited for the study of irreversible processes and the development of methods of dealing with them. We have a spin system which is almost isolated, which is amenable to calculation and which can be subjected to very precise measurement. On the other hand we have the usual thermal bath with its myriad degrees of freedom which can only be vaguely described even for the most perfect of perfect crystals and hardly at all for a liquid. We then have a comparatively weak interaction between the two which is quite precisely known. We can easily set up the spins in a low entropy state and watch them run down due to interaction with the ' lattice ', and we can even assume that the lattice stays in thermal equilibrium all the time. What more could one want? And yet it is still very difficult to get an answer. Such problems clearly demand the use of the density matrix technique and it is used throughout the book. There are very nice illustrations of the power of this method and of ways of dealing with randomly varying Hamiltonians in general.

Professor Abragam is a theoretical physicist and this is a theoretical physicist's book with few concessions to lesser brethren. It has the usual sprinkling of ' it may be shown that ', ' evidently ', and so on which usually conceal several pages of calculation and some of the less easily justified assumptions. A considerable knowledge of both magnetic resonance and theoretical physics is required if the experimental physicist is to read it profitably. On the other hand those interested in theoretical methods will find it an intensely interesting and beautiful application of current methods to this special field. Professor Abragam is perhaps even better known for his contributions to the closely related field of electron spin resonance and it might have been expected that he would first have written on this subject. Those of us interested primarily in nuclear resonance count ourselves fortunate in receiving his attention first. It is clear also, as pointed out in the book, that a very great deal of the matter here can be carried over with little change into electron resonance. Let us hope that in due course he will write a sequel from the point of view of the electron and so enlarge on several of the pregnant asides which he allows himself here.

The book covers all the fields of nuclear resonance and is quite as up-to-date as can be expected in such a rapidly expanding field. Nuclear resonance in condensed phases was discovered in 1946. Andrew's well known book, written in 1954, contained effectively all that was known about nuclear resonance at that time in some two hundred rather small pages; a complete account today would rival the Radiation Laboratory series. The book deals with the theory both classical and quantum mechanical and its relation to experimental methods and results, but not at all with hardware. In spite of the title, Principles, many problems are worked out in considerable detail and useful answers obtained; some are new but most are ones which lie scattered through a vast literature. The relevance of spin dynamics in relation to the generalization of physical ideas is amply illustrated and in particular by the chapter on Spin Temperature. The reality of negative temperatures (which are *hotter* than $+\infty$°K) and of sets of systems at different temperatures in the same space is beautifully established. A hint, and sometimes more, of the applications is given. For instance the determination of molecular structure by high resolution magnetic resonance and of enhanced nuclear polarization and masers by dynamic polarization. The final chapter takes up the most difficult problem of all, ' the power and the glory ' of spin systems in matter tormented by powerful electromagnetic fields which by brute force keep them from their well earned relaxation and force them to reveal yet more of their subtlest and " geistlichste " interactions with the rest of the Universe.

The book is well made in the manner expected of the Oxford University Press. It betrays little evidence of having been written by a person whose native language is not English. The diagrams are clear and the photographs apt; many of the latter were specially prepared for this book.

The quotations under the Chapter headings must not pass without mention since here Professor Abragam's well known sense of humour breaks through. " Liberté, égalité, reversibilité " for the chapter on the motion of free spins is very apt and nicely irreverent—for a Frenchman. This reviewer has to admit to being initially puzzled by " Some Like it Hot " for Spin Temperature and was delighted to be able to see Marilyn Monroe in the interests of his scientific education.

A fine book which will dominate the field for many years.

J. G. Powles.

BOOK REVIEWS

An Introduction to the Theory of Vibrating Systems. By W. G. BICKLEY and A. TALBOT. (Oxford University Press, 1961.) [Pp. xiv + 238.] 30*s.*

Chapter XIV of this book is entitled " Vibrations of Beams and Whirling of Shafts " and no better guide to its scope could be offered. Written by mathematicians for students of engineering the emphasis throughout the greater part of this book is such that the reader is left in no doubt that it is the potential engineer whom the authors are addressing. Its great merit is that it really is what its title proclaims it to be—an introduction. Slightly more than half the book is devoted to an exposition of the behaviour of discrete systems. Energy methods, including Lagrange's equations of motion, Rayleigh's principle, and its application are all discussed in some detail. Continuous systems are introduced by way of two chapters on wave propagation and sound, and if one misses the reference to physical (as opposed to engineering) systems these sections are mercifully free from the arid wastes that the vibrations of membranes, plates and shells so often form in a book of this kind. The final chapter contains a brief treatment of non-linear vibrating systems. One disappointing feature of this book is the nature of the line diagrams. There are eighty of these altogether ; very few have any captions whatever and some are so devoid of labelling or of supplementary information as to be virtually useless. There is a useful set of problems at the end of each chapter.

E. W. LEE.

Nouvelles Propriétés Physiques et Chimiques des Metaux de tres haute Pureté. (C.R.N.S., 1960.) [Pp. 206.]

One of the most striking things about crystalline solids is the way that many of their properties can be changed out of all recognition by the most minute traces of critical impurities. One nitrogen atom amongst one hundred thousand iron atoms can completely change the mechanical behaviour of steel. Copper can be completely embrittled by a minute trace of bismuth. Ultra-pure aluminium recrystallizes at −70°C. The semiconducting properties of silicon and germanium depend crucially on certain impurities present to as little as one atom in one thousand million.

It was therefore exiting when Pfann developed his zone-refining process for the purification of solids and in the last ten years a great deal of research has been done on all the various effects of ultra-refinement on the properties of metals and other solids. A strong Frecnh school has been built up in this field and it is not surprising to see their interest displayed by the publication of this book which presents the contributions made at an international colloquium held at the Centre National de la Recherche Scientifique. This is a book for specialists, i.e. metallurgists and solid-state physicists and chemists, but for these it is most valuable as a record of all the exciting new things that have been discovered in this field in recent years.

A. H. COTTRELL.

Theoretical Physics in the Twentieth Century. A Memorial Volume to Wolfgang Pauli. Edited by A. FIERZ and V. F. WEISSKOPF. (Interscience, 1960.) [Pp. vi + 328.] 72*s.*

This volume, originally planned to celebrate Wolfgang Pauli's 60th birthday, became, through his untimely death, a memorial volume. It is a striking testimonial from thirteen eminent physicists, many of them his former pupils or collaborators, to the remarkable influence he exerted on most branches of theoretical physics during the twentieth century.

The articles, which, incidentally, are not written for the layman, fall into two categories. On the one hand are summaries of branches of physics in which he had an especial interest, and on the other are accounts of the development of quantum mechanics. In this latter field it was not only Pauli's direct contributions that were important but also the very fact of his existence and his continual demand for rigour and clarity from those involved in the development.

In a short review it is impossible to comment on the articles in detail ; suffice it to say that they are all of the high standard to be expected of the various authors concerned.

Quantum mechanics, and in particular the Pauli exclusion principle, is discussed by R. Kronig, W. Heisenberg and B. L. van der Waerden. Articles covering both the early days of quantum field theory and recent developments are contributed by G. Wentzel, F. Villars, R. Jost, and L. D. Landau. R. E. Peierls, and H. G. B. Casimir write on solid state physics ; M. Fierz on statistical mechanics and V. Bargmann on relativity. Finally there is an article by C. S. Wu on the neutrino, a particle first postulated by Pauli. A foreword by Niels Bohr, which sets the tone of the whole book, gives brief autobiographical details and paints an overall picture of Pauli's gigantic contribution to physics. Who can disagree with his final assessment that " . . . he (Pauli) became more and more the very conscience of the community of theoretical physicists " ? R. J. BLIN-STOYLE.

The Fabric of the Heavens. By S. TOULMIN and J. GOODFIELD. (Hutchinson, 1961.) [Pp. 272.] 25*s.*

This is the first of four volumes planned to show the non-scientist the unfolding of scientific thought from ancient times to the present day. It deals with the growth of astronomy (up to a brief glimpse of today's expanding universe) and of the related fundamental concepts of dynamics, such as force, inertia, and energy. It is not written for people who want precise data or dates ; instead it paints, on a broad canvas, that magnificent pageant of brilliant men—from Babylonian astronomers and Greek philosophers to the mathematicians and physicists of today—whose speculations, arguments and observations have taken us where we are today. Stars, comets, lightning and locusts all are things of the sky and were pretty much of a kind to early men though to us they are literally worlds apart. Today it seems quite obvious that a body moves in a straight line indefinitely unless a force acts on it ; but here we are shown why Aristotle, knowing what he did, would conclude that terrestrial bodies were at rest unless pushed while celestial bodies moved in eternal circles, as shown by the work (well documented) of Babylonian astronomers. Again, I for one, was brought up to think that today's view was won in a pitched battle between the schoolmen and those first modern minds, Copernicus, Galileo and Newton. Now we learn that the heliocentric picture was well presented by Aristarchos in Hellenistic times, was discussed and eventually abandoned ; that much of Galileo's achievement was foreshadowed by scholars (some anonymous) who preceded him by a century or two. The style is readable and vivid, and that slow groping for modern concepts throughout history is rubbed in most effectively. The absence of an index is to be deplored. O. R. FRISCH.

Electrostatic Separation of Mixed Granular Solids. By O. C. RALSTON. (Elsevier: D. Van Nostrand, 1961.) [Pp. vii + 262.] 24*s.*

Anybody concerned, even remotely, with the patent literature in this field will find Mr. Ralston's book invaluable. To reduce his subject to order he has had to range from the textbook on electricity to patent specifications of doubtful quality. His success in dealing with the latter clearly stems from an intimate familiarity with the behaviour of electrostatic separators in different industries, and his descriptions are enlivened from time to time by pertinently humorous comment. The reader is aware of being introduced to the technology by one of the few people who have watched the know-how accumulate through the years.

Where basic physics is concerned, however, the author is less sure of himself ; some of the exposition is rather confused, and the book contains at least half a dozen fundamental errors. We are informed, for instance, that a capacitative current leads the e.m.f. by 90° because of dielectric hysteresis. There is a section purporting to summarize Fröhlich's contribution to dielectric theory which, in fact, goes no further than Faraday and Maxwell. It is unfortunate that the description of separators stops short of giving an account of the theories worked out by Schnitzler and by Mora in publications which can only be consulted with difficulty.

Such deficiencies detract from the value of the book but they can be corrected by the wary reader, who will find plenty of value to set against them. When read, the book is still useful for reference ; it is a pity, therefore, that the publishers did not see fit to provide it with a more durable binding. W. R. HARPER.

Radio Waves in the Ionosphere. By K. G. Budden. (Cambridge University Press, 1961.) [Pp. vxiv + 542.] 95*s.*

Ever since the first long distance radio wave transmissions showed that the space immediately surrounding the earth could not be regarded as a simple propagating medium there has been a continuous attempt to explain the very complicated process of electromagnetic wave propagation in the ionosphere. The combined development of both theoretical and experimental work has resulted in a reasonably complete understanding of these processes and Dr. Budden's very comprehensive book contains all the more important theoretical aspects of this work.

The book gives a very complete mathematical picture of propagation and reflection in the ionosphere and consequently requires a reasonably high mathematical standard from the reader, who will also need to know about classical wave propagation and optics.

In Dr. Budden's treatment the ionosphere is assumed to be horizontally stratified, and the problem of propagation is then approached in two ways. In the first half of the book it is assumed that the change in the properties of the medium *per wavelength* are quite small so that the ray theory may be used. For relatively low frequencies (below about 2 Mc/s) this approximation cannot be made and the second half of the book gives the full wave theory that must be used in estimating propagation conditions at these low frequencies. The use of the full wave theory involves some quite severe mathematics

This book is recommended to the intending specialist in this field, but is not for the beginner. I. M. Stephenson.

An Introduction to Celestial Mechanics. By T. E. Sterne. (Interscience Publishers, Inc., New York and London, 1960.) [Pp. 206.] Cloth, 34*s.* ; paper, 19*s.*

The physicists and engineers concerned in developing space vehicles cannot afford to be entirely ignorant of celestial mechanics, and it is primarily to them that Professor Sterne directs this excellent introductory test. Chapter 1 treats the fundamentals of orbits under a central inverse-square-law force. Chapter 2 gives an admirable exposition of the potential of irregular bodies, and, in particular, of nearly spherical ones. Chapter 3 describes the coordinate systems, Chapter 4 the general astronomical perturbation-techniques, and Chapter 6 numerical integration methods. In this part of the book the author is acting as an expert guide over known territory. The lengthy Chapter 5, however, is virtually a monograph on earth-satellite orbital theory, in which Professor Sterne expounds his own novel generalized methods. These methods, although powerful, do not seem to be the most suitable to present in an introductory text, because those dealing with the gravitational field involve difficult concepts, and those dealing with the atmosphere do not carry the analytical solution far enough. With this slight reservation, however, the book can be warmly recommended. It is also well produced, apart from a rather large quota of misprints, and, for a specialized book, comparatively cheap. The book would have been even better if Professor Sterne had taken more note of work done outside the United States; but it would be unfair to single him out for criticism on this score. D. G. King-Hele.

The Technique of Optical Instrument Design. By R. J. Bracey. (English Universities' Press Ltd., 1961.) [Pp. 316.] 35*s.*

It is stated on the dust jacket that " this book was written partly for people who design or have occasion to specify optical parts of instruments and also for the larger number who are interested in the efficient use of such intruments or have occasion to discuss their developed or inherent potentialities ". Most of these people require a concise and well-organised treatment of the subject with the information clearly presented and readily accessible. The information is usually there but the presentation leaves much to be desired. The author's approach to the subject is mainly via the trigonometrical ray-tracing equations and ray aberrations. Inevitably this leads to a rather inelegant treatment, making it difficult to give a clear overall picture of the subject. It is here that the book fails—one cannot see the wood for the trees. However, the book does contain a wealth of detailed information that will be of interest to the student of optical design who already has an overall picture of the nature of the aberrations. It is one of the few books that really discusses design techniques rather than the properties of completed lens systems.

The detailed discussions of individual topics are sound and reliable and the book includes a number of topics seldom treated—lens-mirrors and the chromatic aberration of a thick lens, for example. On the other hand, there is no mention of frequency response in the chapter on image assessment; this, surely, is a subject of great importance to those who " have occasion to specify optical parts ".

R. S. LONGHURST.

Field computations in Engineering and Physics. By A. THOM and C. J. APELT. (Van Nostrand, 1961.) [Pp. vii + 165.] 30*s.*

With the current interest and rapid advances in methods of numerical analysis, any book which deals carefully and authoritatively with some of the many well-established techniques, is to be welcomed. The present text is especially welcomed, not only because it provides a well written account of the method of Squares developed by Professor Thom more than thirty years ago, but because it shows how the method is applied to a wide range of problems and notably to the solution of non-linear hydrodynamic equations. Although as originally developed the method was intended for use with desk machines, it readily lends itself to use with digital computers.

After an introduction, finite difference formulae and solutions of Laplace's equation are treated. Following chapters deal with symmetromorphic figures and free streamline problems, Poisson's equation, convergence and propagation of error in Laplace and Poisson fields, fourth order linear equations, Navier-Stokes equations, flow of a compressible fluid, solution to problems with axial symmetry.

The treatment is thorough, well-illustrated with examples, and the book provides excellent value for money.

C. A. HOGARTH.

Nuclear Sizes. By L. R. B. ELTON. (Oxford University Press, 1961.) [Pp. 114.] 15*s.*

The investigation of the distribution of charge and mass in a nucleus and the understanding of its significance is an important task of the nuclear physicist. The fact that a nucleus seems to have a surprisingly well defined radius enters into all aspects of nuclear physics and conversely a wide variety of experimental techniques have been used to determine radii. These range through high energy electron scattering, meson and nucleon scattering and the interpretation of all these experiments is reviewed in this monograph. In addition such studies throw light on the exact characteristics of the nuclear surface and this subject is also clearly discussed. The emphasis throughout the book is not on the experiments but rather on the insight which they can give us into the nature of nuclear matter. Some knowledge of the theoretical techniques used in nuclear theory is required of the reader. This will be a useful guide book for the nuclear physicist or theorist although it may well turn out to be a progress report since our knowledge is clearly not yet complete.

E. B. PAUL.

The Special Theory of Relativity. By HERBERT DINGLE. (Methuen, London, 1961). (Methuen's Monographs on Physical Subjects.) [Pp. vi + 94.] 8*s.* 6*d.*

This is a somewhat unusual presentation of the theory of relativity. It starts in the conventional way with a discussion of the classical experiments (Fizeau, Michelson-Morley, Kennedy-Thorndike) and presents clearly and in considerable detail the conflict between their results and the idea of an ether. The rest of the book is an elaboration of the author's view that everything falls into place if we assume that not the length l but the quantity $l\surd\{1-(v/c)^2\}$ remains invariant when the same phenomenon is referred to different inertial frames. From this he succeeds in deriving the Lorentz transformations and hence the whole of relativity, and there are a good many shrewd and instructive comments. But the emphasis on length seems arbitrary; surely time measurements, say with an atomic clock, are just as fundamental, and the formal symmetry of space and time in relativity theory contains an element of beauty which remains neglected in this book. As one might expect, there is little comment on the problems and paradoxes of time measurement, and, by the way, the name of Einstein does not occur.

O. R. FRISCH.

Turbulence: Classical Papers on Statistical Theory. Edited By S. K. FRIEDLANDER and LEONARD TOPPER. (New York and London: Interscience Publishers Inc., 1961.) [Pp. ix+187.] Price $6.00 (45*s*.).

In his great work *Hydrodynamics*, Sir Horace Lamb, after devoting many hundreds of pages to the exposition of the ' classical ' theory of the motion of fluids, began the discussion of turbulence by remarking that the reader had now reached the most difficult part of the subject. Difficulties, indeed, confront one from the very start, for there is no simple definition of the turbulent state and perhaps the most that can be said is that when turbulence is present, it can be recognized by finite random oscillations in the velocity of the stream. Such oscillations cause the rapid transfer of momentum, mass and energy from one region to another, so that the study of turbulence is indissolubly linked with that of diffusion.

The present volume is concerned with a method of attack which is now recognized as of major importance. It began, curiously enough, with a contribution to a journal mainly devoted to pure mathematics. In 1921, G. I. Taylor (now Sir Geoffrey Taylor, F.R.S.) turned his attention to a problem which has been known by the unusual name of ' the drunkard's walk ' which, briefly, amounts to the determination of the probable distance from the origin of a particle (or an inebriated citizen) taking, in a given interval of time, a succession of randomly directed steps. (In laboratory physics, such a phenomenon is seen in Brownian motion). Taylor, by introducing the concept of the ' Lagrangian ' correlation between the velocity of a particle at two times, was able to extend the problem to diffusion by continuous movements. His main result was that the standard deviation of the distances travelled by the particle in a given time, and therefore the diffusion of the particles, is related in a simple fashion to the Lagrangian correlation coefficient.

Apart from an application to micrometeorology, little use was made of this work until 1935, when Taylor returned to the subject and in a remarkable series of papers showed that the ideas formulated earlier could be used to produce a more rational theory of turbulence than had existed hitherto, including results capable of experimental verifications in wind tunnels. Not the least significant of Taylor's many concepts was that of the energy spectrum of turbulence. Another notable advance came in 1941 when the Russian mathematician A. N. Kolmogoroff introduced the idea of ' locally isotropic ' turbulence and showed how energy is transferred between eddies of different size.

This book contains the original text of the dozen or so main contributions in the subject, including H. L. Dryden's review of progress until 1943, translations of Kolmogoroff's three classic papers and the work of von Karman, Howarth and Lin. In bringing together in one slim volume the basic material the editors have performed a most useful service. Anyone who wishes to learn about the statistical theory of turbulence could not do better than to acquire this volume, which will always be a valuable work of reference.

O. G. SUTTON.

Physics of the Solar Chromosphere. By R. N. THOMAS and R. G. ATHAY. (New York and London: Interscience Publishers, Inc., 1961.) [Pp. viii+422.] 112*s* 6*d*.

It seems a long time since the sun's chromosphere was regarded as a definite structure or stratum between the site of the dark-lined spectrum and the more mysterious corona familiar at eclipses. We know nowadays that it is not a fixed layer but a region of turbulence and of gradients which differ radically for the frequencies associated with different atomic states. Brief transients in those states give rise to the ionic jets which penetrate the corona and are signalled to us as sporadic radio and optical flares, controlling our aurora, geomagnetic and cosmic ray phenomena.

The present volume, by two members of the famous Colorado institutes for solar and geophysical research, reports some vital contributions to chromospheric physics, based particularly on the eclipse seen from Khartoum in 1952. But the book also contains some excellent systematic progress in the problem of adapting the mathematics of thermodynamic, radiative, and general statistical equilibria to situations dominated by gradients which fluctuate violently. The topic is urgent in solar physics, but is difficult, and considerable education in some more general accounts of the sun and its transients would be desirable before reading these clear but closely-knit chapters.

MARTIN JOHNSON.

SUBJECT INDEX (Vol. II)

Annual exhibition of The Institute of Physics and The Physical Society, 236.
Atomic Energy Establishment, Winfrith: opening, 63.

Conference of the Physical Sciences Section of The Association of Teachers, 71.
Contemporary Reading, 73, 231, 319, 479.
Corso di aggiornamento, 115.

DSIR grants and awards, 240.
DSIR research establishments, 475.

Electrification and contact of solids, 345.
Electromagnetic induction, 286.
Energy for the world's needs, 81.

Faraday and the structure of matter, 93.
Friction of solids, 198.

Gyromagnetic ratio of the proton, 350.

Institute of Physics and the Physical Society: nuclear physics meeting, Glasgow (September 1960), 59.
Interference spectroscopy, 268.
International conference on physics education, 69.
International Scientific Radio Union, 14.

Magnetic dipole and a current-carrying coil, 472.
Masers, 1.
Meteorology, advancing fronts in, 165.
Microtron, the, 277.
Modern physics in schools, 403.

National Chemical Laboratory: open day, 312.
Neutrino and antineutrino losses in nuclear processes, 423.
Neutron booster, the, 15.

Non-linear vibrations, 143.

O.E.E.C. seminars: use of television for teaching of science, 14; mathematics for physicists and engineers, 318.
Optics at Imperial College, 235.

Particles, field theory and force, 325.
Physical Science Study Committee: (3), planning and structure, 39.
Physical science in the training college, 298.
Physics at the British Association (Cardiff meeting, 1960), 55.
Physics for grammar schools, 381.
Physics of diamond, 217.
Plasma, the nature and properties of, 447.
Post Office Research Station, Dollis Hill, London: open days, 157.

Radar astronomy, 116.
Radio emission from interstellar neutral hydrogen, 428.
Research day at The Royal Institution (28 November 1960), 223.

Satellite orbits and their geophysical implications, 253.
Scalpel and spectroscope, 367.
Sceptical physicist, the, 21.
Science in civil defence, 245.
Science in space, 106.
Science Masters' Association: annual meeting, 238.
Single crystals, applications of, 409.

Tandem generator, the, 31.
Thermoelectric power generation, 302.
Transuranic elements, research on, 385.

Voltage reference systems, 463.

X-ray microscopy, 183.

AUTHOR INDEX (Vol. II)

Allen, W. D.,the tandem generator, 31.

Bateman, P. J., thermoelectric power generation, 302.
Bates, L. F., physics at the British Association, Cardiff meeting, 1960, 55.
Birss, R. R., electromagnetic induction, 286.
Blin-Stoyle, R. J., particles, field theory and force, 325.
Braddick, H. J. J., interference spectroscopy, 268.
Brown, G. W., accurate voltage reference systems, 463.

Book Reviews:
- Aharoni, J., The Special Theory of Relativity, 78.
- Benson, O. O., Jr. and Strughold, H. (Eds.), Physics and Medicine of the Atmosphere and Space, 324.
- Bickley, W. G. and Talbot, A., An introduction to the theory of vibrating systems, 484.
- Bracey, R. J., The Technique of Optical Instrument Design, 486.
- Brailsford, F., Magnetic Materials, 242.
- Budden, K. G., Radio Waves in the Ionosphere, 486.
- Buerger, M. J., Crystal Structure Analysis, 243.
- Buxton, C. R. and Jackson, H. S., Translation From Russian for Scientists, 241.
- Clark, R. E. D., Christian Belief and Science, 244.
- De Broglie, L., Non-linear Wave Mechanics, 78.
- Dingle, H., The Special Theory of Relativity, 487.
- Elton, L. R. B., Nuclear Sizes, 487.
- Faraday, Michael, Advice to a Lecturer, 324.
- Fierz, A. and Weisskopf, V. F. (Eds.), Theoretical Physics in the Twentieth Century, 484.
- Friedlander, S. K. and Topper, L. (Eds.), Turbulence: Classical Papers on Statistical Theory, 488.
- Landé, A., From Dualism to Unity in Quantum Physics, 323.
- Lectures in Theoretical Physics, 322.
- Marion, J. B. and Fowler, J. L. (Eds.), Fast Neutron Physics, 323.
- Nelkon, M., General Science Physics, 242.
- Nouvelles propriétés physiques et chimiques des metaux de tres haute pureté, 484.
- Radio Research, 1959, 80.
- Ralston, O. C., Electrostatic Separation of Mixed Granular Solids, 485.
- Ratcliffe, J. A. (Ed.), Physics of the Upper Atmosphere, 79.
- Roberts, J., High Frequency Applications of Ferrites, 244.
- Shire, E. S., Classical Electricity and Magnetism, 321.
- Small-scale Preparation of Fine-grain Photographic Emulsion, 323.
- Smith, C. J., General Properties of Matter, 322.
- Sterne, T. E., An Introduction to Celestial Mechanics, 486.
- Thom, A. and Apelt, C. J., Field Computations in Engineering and Physics, 487.
- Thomas, R. N. and Athay, R. G., Physics of the Solar Chromosphere, 488.
- Tolansky, S., Surface Microscopy, 243.
- Toulmin, S. and Goodfield, J., The Fabric of the Heavens, 485.
- Valentine, J. M., Teach Yourself Atomic Physics, 241.
- West, J. C., Analytical Techniques for Non-linear Control Systems, 242.

Chalmers, J. A., the correspondence between a current-carrying coil and a magnetic dipole, 472.

Davies, R. D., radio emission from interstellar neutral hydrogen, 428.
De Barr, A. E., the applications of single crystals, 409.

Evans, J. V., Radar astronomy, 116.

Essay Reviews:
- Abragam, A., The Principles of Nuclear Magnetism. J. G. Powles, 482.
- Stoops, R. (Ed.), La Structure et l'Evolution de l'Univers. E. H. Linfoot, 75.

Freitag, E. H., the friction of solids, 198.

Harper, W. R., electrification following the contact of solids, 345.

Jennings, R. E., The microtron, 277.

King-Hele, D. G., satellite orbits and their geophysical implications, 253.

Lee, E. W., non-linear vibrations, 143.
Lenihan, J. M. A., the scalpel and the spectroscope, 367.
Lewis, G. M., neutrino and antineutrino losses in nuclear processes, 423.

McAulay, J., science in civil defence, 245.
Mitchell, E. W. J., the physics of diamond, 217.

Nixon, W. C., X-ray microscopy, 183.

Poole, M. J., the neutron booster, 15.

Rotherham, L., energy for the world's needs, 81.

Saunders, V. T., opening of the Atomic Energy Establishment, Winfrith, 63; Post Office Research Station, Dollis Hill, London: open days, 157; the National Chemical Laboratory: open day, 312.
Shoenberg, D., low temperature physics, 223.
Sister St. Joan of Arc, the teaching of modern physics in schools, 403.
Stephenson, I. M., science in space, 106.
Stevens, K. W. H., masers, 1.
Sutton, Sir Graham, C.B.E., advancing fronts in meteorology, 165.

Tricker, R. A. R., the sceptical physicist, 21.

Vigoureux, P., the gyromagnetic ratio of the proton, 360.
Von Engel, A., the nature and properties of a plasma, 447.

Author Index (*Vol. II*)

WALDRON, M. B., a survey of research on the transuranic elements, 385.

WENHAM, E. J., conference of the Physical Sciences Section of the Association of Teachers and departments of education, 71; physical science in the training college, 298.

WHITE, S., Physical Science Study Committee: (3), the planning and structure of the course, 39.

WILLIAMS, L. P., Faraday and the structure of matter, 93.

WILLIAMS, W. S. C., the nuclear physics meeting of the Institute of Physics and the Physical Society, Glasgow, September 1960, 59.

Printed by WARREN AND SON, LTD.
(a subsidiary of Taylor and Francis, Ltd.)
The Wykeham Press, Winchester
England.